Ciencias de la Computación
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE

REVISTA DEL DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN DE LA UNIVERSIDAD DE CHILE

Bits DE CIENCIA

EDICIÓN N°21 | AÑO 2021

INTELIGENCIA ARTIFICIAL

Hacia una política chilena de inteligencia artificial, nacida en contexto de pandemia
/ Andrea Rodríguez

Sistemas de toma de decisiones automatizadas: ¿De qué hablamos cuando hablamos de transparencia y del derecho a una explicación?
/ Catherine Muñoz, Jeanna Matthews y Jorge Pérez

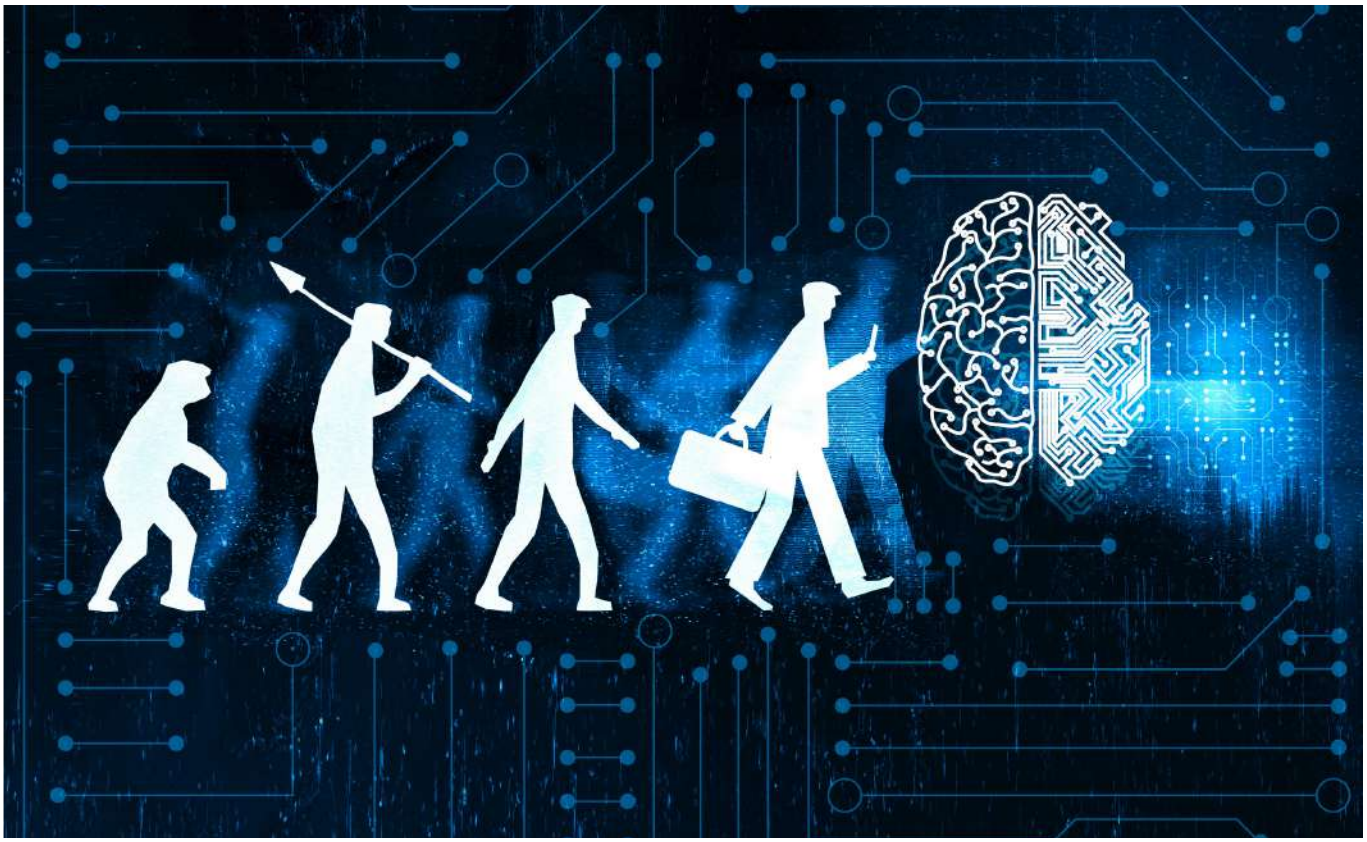
Aplicaciones de la IA: Colección de artículos independientes abordando distintas aplicaciones de la inteligencia artificial
/ Varios autores

Contenidos

Revista Bits de Ciencia Nº21
ISSN 0718-8005 (versión impresa)
www.dcc.uchile.cl/revista
ISSN 0717-8013 (versión en línea)

Departamento de Ciencias de la Computación
Avda. Beauchef 851, 3º piso, edificio norte. Santiago, Chile. 837-0459 Santiago

www.dcc.uchile.cl
56 22 9780652
revista@dcc.uchile.cl
/dccuchile



Editorial

FEDERICO OLMEDO
Editor General
Revista Bits de Ciencia

Impulsada principalmente por los avances del aprendizaje automático (*machine learning*), la Inteligencia Artificial (IA) ha adquirido un rol prominente en los últimos años. Son cada día más los actores, tanto privados como públicos, que la están incorporando para mejorar sus procesos o automatizar tareas que tradicionalmente requerían la intervención de un ser humano. El impacto que esta tecnología está teniendo en la sociedad, particularmente en la toma de decisiones, es innegable: desde determinar qué películas o series nos recomienda nuestra plataforma de *streaming* favorita, hasta decidir quién es apto/a para la otorgación de un crédito o quién es el/la "mejor" candidato/a para un puesto de trabajo.

En este número de la Revista hemos decidido abordar, por tanto, algunos de los aspectos fundamentales de esta tecnología. Hacemos un repaso de su desarrollo histórico, describimos algunas de sus aplicaciones, y discutimos los desafíos éticos —y también paradojas— que conllevan su aplicación. A nivel nacional, presentamos tres iniciativas institucionales recientes, gestadas en torno a la IA, y describimos en mayor profundidad la iniciativa gubernamental para promover y regular la IA, la Política Nacional de Inteligencia Artificial.

Finalmente abordamos otros temas atingentes y de actualidad en las secciones de Investigación Destacada (*Prediciendo indicadores en el retail*), Premio Turing (*La revolución de la animación 3D por computadora*), Computación y Sociedad (*Juan Álvarez Rubio: A medio siglo de mi encuentro con la computación en la "Escuela de Injeniería"*), y Doctorados.

Espero que disfruten de este número extendido (2021) de la Revista. Cualquier comentario o sugerencia, no duden en escribirnos a revista@dcc.uchile.cl. ▪

Prediciendo indicadores en el retail

NELSON BALOIAN

Profesor Asociado del Departamento de Ciencias de la Computación de la Universidad de Chile, y profesor visitante regular en las universidades de Waseda, Japón, y de Duisburg-Essen, Alemania. PhD en Ciencias por la Universidad de Duisburg, Alemania. Sus áreas de interés de investigación han sido los sistemas de apoyo computacionales para el aprendizaje, sistemas distribuidos y machine learning. Ha sido autor de más de 30 artículos en revistas indexadas y más de 100 en conferencias internacionales.

nbaloian@dcc.uchile.cl

JOSÉ A. PINO

Profesor Titular del Departamento de Ciencias de la Computación de la Universidad de Chile. Cofundador del DCC, ha servido como Presidente de la Sociedad Chilena de Ciencias de la Computación y Presidente de CLEI. Sus áreas de interés actuales son aprendizaje de máquina y administración de procesos de negocio. Su investigación se ha publicado en *journals*, incluyendo Expert Systems with Applications, Information Systems Frontiers y ACM Computing Surveys.

jpino@dcc.uchile.cl

BELISARIO PANAY

Ingeniero Civil en Computación y Magíster en Ciencias mención Computación de la Universidad de Chile. Líneas de trabajo: aprendizaje automático. En Twitter lo encuentras como @belisariops.

bpanay@dcc.uchile.cl

Las tiendas no-online en nuestro país han sufrido fuerte con la pandemia del COVID-19. Como han debido permanecer cerradas durante largos periodos, en éstos no han recibido ingresos, pero sí han debido afrontar sus gastos fijos. En consecuencia, es importante para ellas optimizar su operación cuando las condiciones mejoren. ¿Cómo podría un administrador conocer su situación en cuanto a ventas?

Un primer indicador de la eficacia de las ventas es averiguar la cantidad de personas que visita la tienda por unidad de tiempo. En caso de que nadie entre a la tienda, mal puede haber ventas. Por el contrario, si muchas personas van a la tienda, hay mayor probabilidad de que compren. Entonces, el número de personas que entra a la tienda por unidad de tiempo (*foot traffic*) es un indicador de la eficacia. Pero podría darse que muchas personas visiten la tienda, pero pocas compren, así es que la proporción de personas que compran con respecto al total de visitantes (*conversion rate*) es un segundo indicador apropiado. Un tercer indicador es el total de ventas en dinero realizadas por unidad de tiempo.

Estos indicadores no sólo pueden servir para conocer la situación pasada, sino que pueden usarse para predecir el desempeño futuro. En la medida que eso sea posible, el administrador puede prepararse para esa futura demanda. Así, puede preparar el número suficiente de vendedores y cajeros, el stock de productos a vender y el flujo de caja esperado.

En este artículo presentamos un proyecto de desarrollo de software de predicción de indicadores para tiendas del retail llevado a cabo por nuestro grupo MARAL (Machine learning Research Applied Lab) con financiamiento de Corfo Innova y apoyo de la empresa Follow Up. El grupo está compuesto, además de los autores del artículo, por Sergio Peñafiel (estudiante de Magíster, ya graduado), Jonathan Frez (candidato a Doctorado) y Cristóbal Fuenzalida (estudiante de Ingeniería Civil en Computación).

El problema

Un enfoque al tema de la predicción de indicadores es usar información previa sobre estos mismos indicadores, pero ¿cómo obtenerla? Aquí viene el aporte de la empresa Follow Up. Esta compañía

nacional ha instalado cámaras a la entrada de gran número de tiendas de Chile, Colombia, Perú, Japón y otros países. Con ayuda de ellas, ha construido una base de datos del *foot traffic* de esas tiendas: la cantidad de visitantes por hora. Los datos de los otros indicadores se pueden obtener del software de ventas de las tiendas mismas.

La predicción que se pretende debe cumplir dos requerimientos. Por una parte, el software debe proveer la importancia que cada variable de entrada tiene en la resultante salida. Esto porque sabiendo ese peso, los administradores pueden reaccionar mejor a cambios inesperados de las condiciones ambientales, tales como un nuevo día festivo, o una campaña de marketing. El segundo requerimiento se refiere al grado de confianza o por el contrario, de incertidumbre, de la predicción. Estos requerimientos implican que el modelo a usarse debe tener algún grado de transparencia, en oposición a los modelos de "caja negra", que no entregan ninguna información adicional a la predicción misma. Normalmente, los modelos de caja negra entregan predicciones con menos error que los métodos transparentes, por lo que típicamente se paga un precio en términos de calidad de predicción si se quiere transparencia. En nuestra investigación, entonces, intentamos desarrollar un método de tipo transparente que tuviera similar calidad de predicción que los modelos de caja negra.

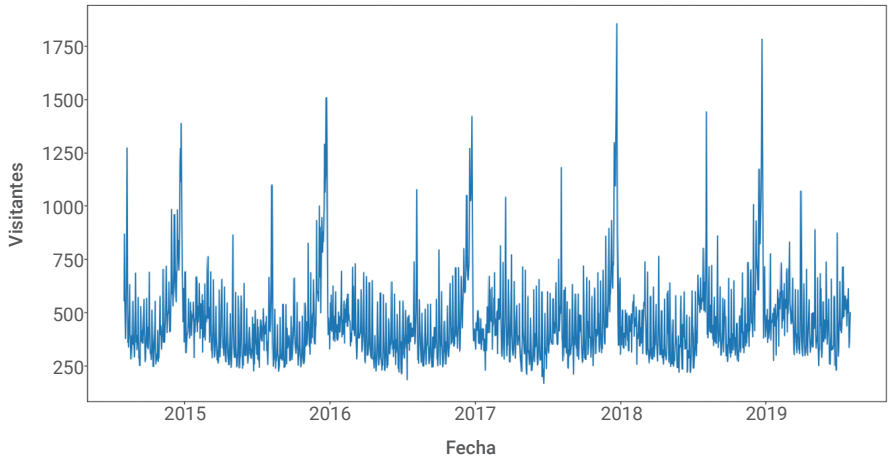
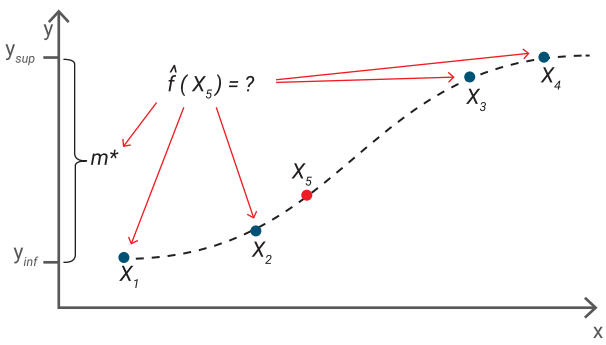


Figura 1. Número de visitantes a una tienda por hora de acuerdo a la fecha.



Observación	Masa
X_1	0.1
X_2	0.4
X_3	0.2
X_4	0.1
m^*	0.2

Figura 2. Ilustración del método.

Metodología

El método usado para encontrar un modelo satisfactorio fue el siguiente. Después de investigar los modelos existentes, se trató de mejorar alguno promisorio. En seguida, se trabajó en el *embedding,* es decir, investigar la mejor manera de codificar la información disponible (en un vector) para que sea entendida por el modelo. Posteriormente, se realizaron experimentos con el modelo estudiando cómo se desempeña con nuestros datos. En cada iteración, se cambiaba el *embedding* o el modelo, hasta obtener resultados satisfactorios.

Los datos utilizados en nuestro estudio no fueron todos los disponibles. Se buscaron datos de tiendas con datos completos durante cuatro años. Sólo 20 tiendas cumplían estos requisitos, con información entre agosto de 2015 y el mismo mes de 2019. La Figura 1 muestra un típico patrón de comportamiento del *foot traffic* en dicho período.

El modelo desarrollado

Después de revisar varios modelos de predicción publicados en la literatura, seleccionamos para mayor investigación el modelo de *Evidential Regression* (EVREG), basado en una extensión difusa de las funciones de creencia, desarrollado por Petit-Renaud y Denoeux, publicado en 2004 [1]. Las funciones de creencia, a su vez, son parte de la Teoría de Dempster-Shafer también conocida como la teoría de evidencia [2]. Este método predice un valor usando un

[Nuestro enfoque] logra un buen desempeño en general, obteniendo resultados comparables a los mejores métodos probados en la literatura.

conjunto de observaciones pasadas. Al predecir, a cada observación se le asigna una masa que representa la similitud con el vector que se va a predecir. Luego estas masas son transformadas en una distribución de probabilidades con la que se calcula un valor esperado. En términos simples podemos explicarlos con la Figura 2.

Acá tenemos un ejemplo donde se tiene un conjunto de observaciones con 4 vectores (en azul). Éste sería el conjunto de entrenamiento de un problema. Y lo que se necesita es predecir la salida de una nueva observación (en rojo). Para encontrar el valor estimado de salida de esta nueva observación, se puede suponer que un vector que esté cerca de ella va a tener una salida parecida. En este ejemplo el segundo vector es el más cercano, así que se puede suponer que éste tiene la salida más parecida. Para reflejar esto, se le asigna una "masa" o importancia a cada uno de los puntos en el conjunto de observaciones según una función de distancia (por ejemplo, la euclidiana). Mientras mayor es la masa más importante es el punto. Esto significa que mientras más cerca está un punto de la nueva observación, más importante es la evidencia que éste entrega. Lo anterior, es similar a un regresor de *k*-vecinos más cercanos donde se calcula una contribución según una distancia. Pero la teoría de la evidencia fue creada para razonar con incerteza. Por ejemplo, puede ocurrir que los puntos del conjunto de entrenamiento están muy lejos de la nueva observación, por lo que no se puede estar seguro de la respuesta que se entrega. Para esto se agrega otra fuente de información, en este ejemplo además de las observaciones de entrenamiento se conocen cuáles son los valores por los que se mueve la salida de estas observaciones. Para reflejar la incertidumbre de la respuesta, *Evidential Regression* entrega una masa a esta observación del intervalo de salida, la cual representa el grado de incertidumbre del proceso.

Evidential Regression calcula las masas de cada una de las observaciones en el set de entrenamiento con una distancia usando un vector de características. Este vector de características codifica la información de un problema, por ejemplo, para este problema las dimensiones del vector tenían datos como el trimestre, mes, día del mes y día de la semana del evento que se quería predecir. Al calcular la distancia entre estos vectores de características *Evidential Regression* supone que todas las dimensiones o características de los vectores de entrenamiento son igual de importantes para calcular la similitud o importancia entre vectores. ¿Pero qué pasa si estamos prediciendo *foot traffic* para una tienda que se encuentra cerca de oficinas? Se esperaría que los días de semana afecten mucho la predicción de sus entradas, ya que es en días hábiles cuando más público se observa. Entonces una diferencia de una dimensión como el día de la semana no afectará de la misma manera que una diferencia en una dimensión como el trimestre en que se encuentra el evento. Entonces lo que proponemos es una versión mejorada de *Evidential Regression* que llamaremos *Weighted Evidential Regression* (WEVREG) que usa una distancia ponderada, para el cálculo de masas [3]. Estos pesos serán aprendidos durante la fase de entrenamiento del algoritmo usando descenso de gradiente. Cada peso representa la importancia de la dimensión del vector de entrada para predecir las salidas del modelo y así ayudará al modelo a aumentar su poder de predicción. Además, para disminuir la complejidad del algoritmo, se usará un *k*-NN (*k-nearest neighbors*) donde sólo se calcularán las masas de los *k* vecinos más cercanos.

Método	RMSE
RF	0.1041 ± 0.01
WEVREG	0.1088 ± 0.01
SVM	0.1133 ± 0.01
GP	0.1321 ± 0.01
LSTM	0.1422 ± 0.02
SARIMA	0.1489 ± 0.03

Figura 3. Resultados.

Evaluación

Los datos consisten en datos de tres indicadores *foot traffic*, *conversion rate* y total de ventas de 20 tiendas, entre agosto de 2015 y el mismo mes de 2019. La meta de este problema es predecir cada uno de estos indicadores para el mes de julio de 2019, usando todos los datos que se tienen disponibles. Para esto se tenían datos diarios de cada uno de estos indicadores. El vector de características se creó desagregando el tiempo de los eventos y se les aplicó una representación circular, por ejemplo, el día de la semana (representado con un entero que va de 0 a 6) se separa en el seno y coseno del día de la semana. A esto se agregó una secuencia que tenía los valores de salida de los últimos 6 meses con un intervalo de 1 mes, esto significa que para el día 1 de agosto, se tenían los valores de salida del 1 de julio, 1 de junio y así sucesivamente.

Para evaluar cuantitativamente el método propuesto, se le comparó con otros métodos que han sido usados con anterioridad en la literatura. Estos son métodos como *Random Forest (RF), Long Term Short Memory* (LSTM), *Support*

Otra característica destacable [de nuestro enfoque] es su interpretabilidad [...]: el modelo da los pesos de los atributos después del entrenamiento.

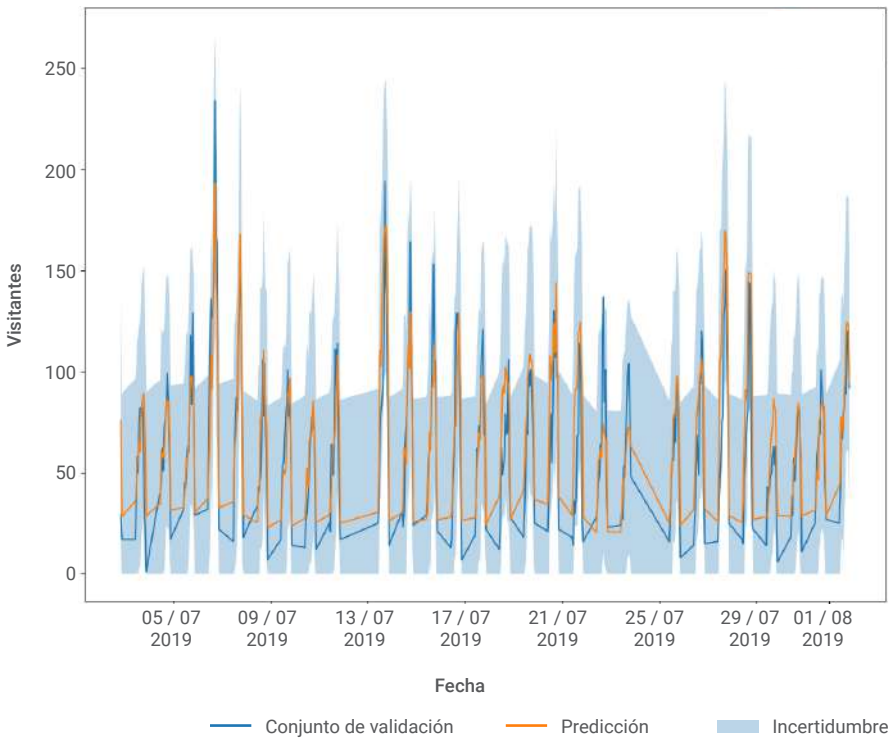


Figura 4. Predicción del *foot traffic*.

Vector Machine (SVM), *Gaussian Process* (GP) y *Seasonal Autoregressive Integrated Moving Average* (SARIMA). En la Figura 3 pueden observarse los resultados. El error medido fue el *Root Mean Squared Error* (RMSE), mientras menor es este error mejor es el método. De todos los métodos puestos a prueba el RF fue el que obtuvo los mejores resultados, pero fue seguido de cerca por nuestro método propuesto, el cual obtuvo el segundo lugar para nuestro conjunto de datos.

En la Figura 4 podemos observar la predicción del *foot traffic* para una tienda en particular. En color azul se ve la curva real (determinada a partir de las observaciones de las cámaras) y en naranjo se ve la predicción de nuestro modelo. Como se puede ver, se ajusta bastante a los datos reales y además entrega un intervalo de variación el cual es la incertidumbre de la respuesta. Las predicciones son calculadas como un valor esperado, cuando se predice una nueva observación, por ejemplo la del día 5 de julio, se calculan las masas (o importancias) de cada uno de sus *k* vecinos. Para llegar a una predicción se multiplican estas masas por los valores de salida de sus vecinos y, además, se agrega el término de incertidumbre el cual se calcula como la masa de la incertidumbre por el valor medio del intervalo en que se mueven los vecinos de la nueva observación. Esto se muestra en la Ecuación (1), donde *N* es la cantidad total de vecinos, *x* es la nueva observación, m_i es la masa del vecino *i*, y_i es el valor de salida del vecino *i*, *m** es la masa de la incertidumbre, *sup y* y *inf y* son el valor máximo y mínimo de salida del conjunto de vecinos.

$$\widehat{y} = \sum_{i=1}^{N} m_i(x) \cdot y_i + \frac{m^*(x) \cdot (\sup_{y \in \mathcal{L}} y + \inf_{y \in \mathcal{L}} y)}{2} \quad (1)$$

Con esto se puede calcular un límite superior e inferior de esta predicción como se muestra en las Ecuaciones 2 y 3.

$$\widehat{y}^* = \sum_{i=1}^{N} m_i(x) \cdot y_i + m^*(x) \cdot \sup_{y \in \mathcal{L}} y \quad (2)$$

$$\widehat{y}_* = \sum_{i=1}^{N} m_i(x) \cdot y_i + m^*(x) \cdot \inf_{y \in \mathcal{L}} y \quad (3)$$

Además de entregar una predicción y su intervalo de incertidumbre, el método es capaz de entregar una medida que representa la importancia de cada una de las variables de la entrada que es usada para predecir. En la Figura 5 se puede observar la importancia de cada una de estas variables. Como se puede ver, para esta tienda en particular el valor más importante para la predicción es la secuencia de valores anteriores, en especial, los valores del *foot traffic* registrados hace 1 mes y 6 meses antes del día que se quiere predecir.

Conclusiones

Del análisis de los resultados presentados en la sección anterior podemos ver que nuestro enfoque logra predecir correctamente los indicadores claves del retail. En efecto, en la Figura 4, podemos ver que en general las predicciones están bien

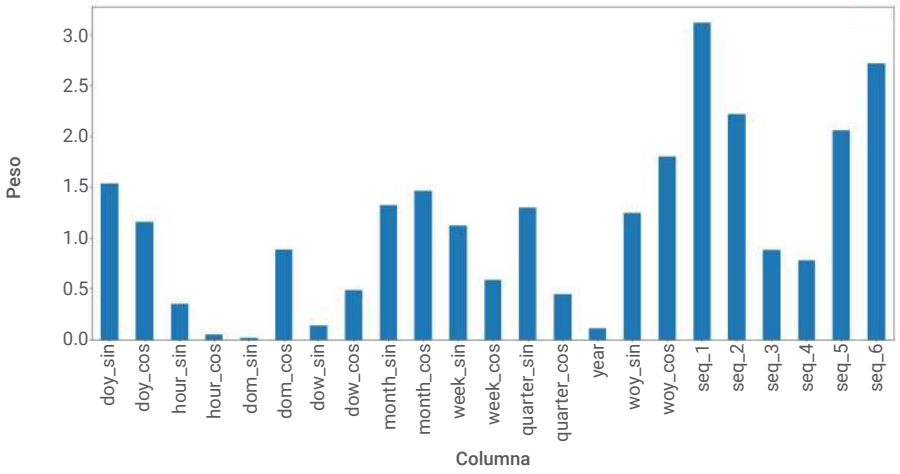


Figura 5. Importancia de los parámetros.

ajustadas a las curvas reales. El método no tiene problemas para detectar las puntas y valles de los valores reales, aunque no alcanza los mismos valores superiores e inferiores. En particular, el modelo apenas alcanza los valores extremos de la predicción. Esta deficiencia puede explicarse por la naturaleza de la predicción con *k*-NN; hay que tener en cuenta que para predecir un valor 0 (el mínimo en nuestro caso), el modelo requiere que todos los vecinos que observa también deben tener el valor 0; si alguno de ellos no tiene un valor 0 entonces "mueve" la predicción hacia el centro.

Otra característica de WEVREG es su capacidad de proporcionar intervalos de variación. En la misma figura, podemos observar que casi todos los valores reales están dentro del intervalo de variación. Sin embargo, podemos observar que el ancho de este intervalo es amplio, cubriendo alrededor del 30% del rango de predicción. Esto podría deberse a que los vectores utilizados por los métodos posteriores al enfoque *k*-NN para cada predicción no son muy similares entre sí, obteniendo una alta incertidumbre para el proceso.

Otra característica destacable del modelo WEVREG es su interpretabilidad. Como se muestra en la Figura 5, el modelo da los pesos de los atributos después del entrenamiento. En este caso, el modelo está utilizando un *embedding* cíclico junto con la secuencia de los seis meses anteriores para una tienda en particular. A partir de esta figura, queda claro que para esta tienda en particular algunos componentes como el año o componentes parciales de la hora y el día del mes no son realmente importantes para predecir sus visitantes. Además, las características más importantes para predecir el número de visitantes parecen ser el número de visitantes observados en meses anteriores. El mes anterior es el más importante, y posteriormente se observa una disminución de importancia seguida de un aumento en el quinto y sexto mes que podría deberse al comportamiento cíclico de los visitantes de esta tienda en particular.

Hablando de desempeño, WEVREG logra un buen desempeño en general, obteniendo resultados comparables a los mejores métodos probados en la literatura [4]. Como se muestra en las Figura 3, RF obtiene el mejor desempeño y este resultado coincide con lo reportado previamente en la literatura [5].

Como se trata de problemas de series de tiempo, se esperaba que LSTM, que es un método basado en aprendizaje profundo, obtuviera los mejores resultados, pero no pudo superar nuestro método propuesto en general. Una posible razón detrás de este bajo rendimiento podría ser el uso de una arquitectura de red única para todas las tiendas. ■

REFERENCIAS

[1] S. Petit-Renaud, T. Denœux: Nonparametric regression analysis of uncertain and imprecise data using belief functions. *Int. J. Approx. Reason.* 35, 2004, 1–28.

[2] G. Shafer: Dempster's rule of combination. *Int. J. Approx. Reason.* 79, 2016, 26–40.

[3] B. Panay, N. Baloian, J.A. Pino, S. Peñafiel, H. Sanson, N. Bersano: Feature selection for health care costs prediction using Weighted Evidential Regression. *Sensors* 20(16), 2020, 4392.

[4] B. Panay, N. Baloian, J.A. Pino, S. Peñafiel, J. Frez, C. Fuenzalida, H. Sanson: Forecasting key retail performance indicators using interpretable regression. *Sensors* 21(5), 2021, 1874.

[5] S. Abrishami, P. Kumar, W. Nienaber: Smart stores: A scalable foot traffic collection and prediction system. In Industrial Conference on Data Mining; Springer: Cham, Switzerland, 2017, 107–121.

Premio Turing 2019:

La revolución de la animación 3D por computadora

Ganadores del Premio Turing 2019.

Ed Catmull

Patrick M. Hanrahan

Fuente: Ed Catmull https://awards.acm.org/award-winners/catmull_1244219 | Patrick Hanrahan: https://news.stanford.edu/2020/03/18/pat-hanrahan-wins-turing-award/

BENJAMÍN BUSTOS

Profesor Titular del Departamento de Ciencias de la Computación, Universidad de Chile. Investigador Asociado del Instituto Milenio Fundamentos de los Datos. Doctor en Ciencias Naturales por la Universidad de Konstanz, Alemania. Líneas de investigación: recuperación de información multimedia basada en contenido, búsqueda por similitud y bases de datos multimedia.

bebustos@dcc.uchile.cl

NANCY HITSCHFELD KAHLER

Profesora Asociada del Departamento de Ciencias de la Computación, Universidad de Chile. Miembro del CEnter for Modern Computational ENgineering (CEMCEN). Doctora en Technischen Wissenschaften por la ETH-Zurich, Suiza. Líneas de investigación: mallas de polígonos y poliedros, algoritmos paralelos (computación en GPU), algoritmos en ciencia e ingeniería computacional y educación en computación. Participa de comisiones y actividades para atraer mujeres a STEM.

nancy@dcc.uchile.cl

Este reconocimiento fue otorgado en el año 2019 a Edwin E. Catmull y Patrick M. Hanrahan, por sus contribuciones en el área de *Computer-Generated Imagery* (imágenes generadas por computadora o CGI). Estas contribuciones han liderado los avances en el área de la computación gráfica, innovando en sus conceptos fundamentales, algoritmos, hardware y software. Es interesante destacar que este Premio Turing 2019 lo comparten dos personas con perfiles bastante distintos. Por una parte, Catmull es un emprendedor e innovador que tuvo un gran impacto en la industria del cine, pero su producción científica tiene relativamente pocas citas. En cambio, Hanrahan es un académico más tradicional, con alto impacto en investigación medido en citas. El Premio Turing 2019 reconoce tanto los aportes de Catmull como de Hanrahan desde sus respectivas perspectivas, dado su innegable impacto en el desarrollo de la computación gráfica en 3D y en CGI.

En las siguientes secciones de este artículo introducimos primero en qué consiste *Computer-Generated Imagery*, luego describimos quién es cada uno de los galardonados y sus contribuciones más importantes, y finalmente concluimos con el impacto que han tenido sus contribuciones no sólo en el área de la computación gráfica, sino también en el cine y en otras áreas de la ciencia e ingeniería.

Imágenes generadas por computadora

El concepto de imágenes generadas por computadora o CGI (por la sigla en inglés de *Computer-Generated Imagery*) se refiere a la generación de imágenes y gráficos 3D aplicadas al arte, cine, televisión, etc., mediante el uso de computadores utilizando algoritmos y técnicas de computación gráfica. La ventaja de utilizar CGI es que permite a los creadores tener una libertad virtualmente ilimitada para generar escenas, incluso "rompiendo" las leyes de la física, permitiéndoles crear escenas que serían muy difíciles o muy costosas de recrear en un escenario real.

Los inicios de la CGI se remontan a las primeras décadas de desarrollo de la computación como ciencia. Ya a finales de la década de los cincuenta, en el film *Vértigo* de Alfred Hitchcock se utilizó CGI para generar una animación en 2D que corresponde a la secuencia de apertura de esta película. En la década de los setenta se empezó a utilizar CGI para generar, en forma rudimentaria aún, pequeñas escenas de acción 2D en películas, y también se empezó a utilizar animación en 3D. Luego, en los años ochenta la película *TRON* fue una de las primeras en hacer uso intensivo de CGI para generar secuencias largas de animación en 3D, de varios minutos de duración. Finalmente, la primera película que fue 100% generada utilizando CGI fue *Toy Story* en 1995 (ver Figura 1). De ahí en adelante, los avances en las técnicas de CGI y en la capacidad de cómputo de los computadores actuales, han permitido el ir generando efectos cada vez más espectaculares y realistas.

El propósito de la animación en 3D es generar escenarios en tres dimensiones con el uso del computador. Para lograr esto, primero es necesario definir cómo se representará la información 3D en el computador. Una de las principales formas para hacer esto es representando los objetos como mallas de polígonos (triángulos o cuadriláteros), en donde cada polígono de la malla se define por las coordenadas en el espacio cartesiano tridimensional de sus vértices. Al colocar polígonos en forma adyacente se pueden formar las superficies de los objetos que se están modelando. Adicionalmente, se puede representar la orientación de la superficie (hacia el interior o hacia el exterior) usando el orden en que se almacenan los vértices de cada polígono. La ventaja que tiene

Figura 1. Cuarta parte de la saga *Toy Story*, cuyas películas de animación 3D fueron totalmente producidas por computadora (CGI).

Figura 2. Ejemplo de objeto 3D representado como malla de polígonos.

el usar mallas de polígonos es que incluso con pocos polígonos se pueden modelar objetos complejos, lo que permite tener una representación simple del objeto y que no ocupa mucha memoria en el computador. En caso que se requiera mayor precisión o nivel de resolución, siempre es posible refinar las mallas y agregar más polígonos para definir la superficie del objeto con mayor nivel de detalle. La Figura 2 muestra un ejemplo de una malla de polígonos que representa un objeto 3D.

Las primeras animaciones generadas usando CGI utilizaron mallas de triángulos para representar objetos simples. Un ejemplo notable es una animación computarizada en 3D de una mano creada por Edwin E. Catmull y Fred Parke en 1972 en la Universidad de Utah. Para este proyecto, Catmull creó un modelo de yeso de su propia mano. Luego, junto a Parke midieron y calcularon manualmente una triangulación 3D del modelo de yeso, usando un par de cientos de triángulos. Finalmente, introdujeron toda esta información en un computador, con lo que produjeron una visualización en 3D de su mano. En el video titulado "A Computer Animated Hand" se observa el modelo 3D de la mano, que puede rotar y flexionar los dedos.[1] Este video ha sido descrito como "revolucionario" para su época, y fundó las bases para todo el desarrollo posterior de la CGI [1].

Edwin E. Catmull

Edwin Catmull fue cofundador de Pixar Animation Studios y presidente de Pixar y Walt Disney Animation Studios. Obtuvo el grado de Bachelor of Science en Física y Ciencias de la Computación (1970) y el PhD en Ciencias de la Computación (1974) en la Universidad de Utah. Durante su carrera fue vicepresidente de la División de Computación en Lucasfilm Ltd., donde dirigió el desarrollo en áreas de computación gráfica, edición de videos, videojuegos y audio digital.

Su motivación por crear películas nació desde muy pequeño inspirado por las películas de Walt Disney como *Peter Pan* y *Pinocho*. Él creó animaciones armando cuadernillos de imágenes, en que página a página contenían dibujos que varían gradualmente. Al mostrar rápidamente las páginas consecutivas, las imágenes parecían animarse simulando un movimiento.

Durante su años en la universidad, realizó dos aportes fundamentales a la

1 | https://vimeo.com/59434349.

Figura 3. La misma malla del objeto 3D de la Figura 2, pero con texturas aplicadas sobre la malla.

Este Premio Turing 2019 lo comparten dos personas con perfiles bastante distintos [...]: Catmull es un [...] innovador que tuvo un gran impacto en la industria del cine [...] En cambio, Hanrahan es un académico [...] con alto impacto en investigación.

computación gráfica: (i) mapeo de texturas (*texture mapping*) y (ii) parches bicúbicos (*bicubic patches*).

Mapeo de texturas. Una textura es una imagen que contiene la forma en que queremos pintar una parte de un objeto, difícil de especificar geométricamente, como por ejemplo la rugosidad de la piel de la mano mencionada en la sección anterior. Aplicado a nuestro ejemplo, el mapeo de texturas permite hacer calzar una imagen real de la piel de una mano, en dos dimensiones, a cada triángulo de la malla que la representa, una vez que el modelo de la mano ha sido proyectado a dos dimensiones para generar su imagen en el computador. El mapeo se realiza desde la textura al triángulo proyectado, en donde los colores de la textura son usados para pintar los píxeles que están contenidos en el triángulo a pintar (ver Figura 3). El aporte de Edmund Catmull fue diseñar un nuevo algoritmo, para hacer calzar una textura con la proyección de un triángulo, todo como parte del mismo proceso de generar la imagen del objeto a pintar [2]. Hasta ese momento, el proceso de mapear texturas era impreciso y lento pues consistía en el proceso inverso: dado un píxel a pintar, se buscaba en el espacio en tres dimensiones, qué parte de la textura asociada a un triángulo proyectado, le correspondía.

Parches bicúbicos. Los *bicubic patches* pueden ser vistos como polígonos de cuatro lados, en que cada lado está representado por un polinomio de grado 3 (curva cúbica). Cada lado necesita dos puntos de control adicionales y en el interior del polígono se requieren cuatro puntos adicionales, para permitir representar superficies curvas de forma más realista que polígonos planares. Catmull introdujo técnicas innovadoras para crear y pintar de manera realista *bicubic patches* en vez de polígonos planos, de las cuales surgió, junto al mapeo de texturas, la técnica del *z-buffer* (descrito al mismo tiempo por Wolfgang Strasser). Para determinar qué partes de la escena modelada (por ejemplo, qué triángulos de la malla que representa la mano mencionada más arriba) serán visibles e influyen en cómo se pinta en la imagen generada, la técnica del *z-buffer* permite seleccionar, al estar pintando la imagen, el color asociado al triángulo más cercano a la cámara (punto desde donde se mira la escena). A cada triángulo de la escena, se le aplican las transformaciones de movimiento y proyección para llevarlo a una representación normalizada, representada por un cubo (-1,-1,-1) y (1,1,1), en donde la imagen a generar coincide con el rectángulo (-1,-1,-1) y (1,1,-1). Este rectángulo se discretiza en píxeles, considerando la resolución que se desea de la imagen a generar. Cada triángulo que queda incluido dentro de este cubo se recorre en función de los píxeles (*scanline*) que cubre, y el color o textura asociada a este triángulo define como pintar estos píxeles. Cada vez que se pinta un píxel se almacena la profundidad en el eje *z* (dentro de este cubo) del triángulo que definió su color. Si aparece otro triángulo más cercano a la cara del cubo que representa la imagen a generar, se usa el color/textura de este nuevo triángulo y se recuerda esta profundidad. Esta técnica almacena, en todo momento, el color y la profundidad del triángulo que está definiendo el color actual. Cuando se terminan de recorrer los triángulos de la escena, se tiene la imagen calculada.

El algoritmo del *z-buffer* fue más tarde generalizado al *A-buffer* para permitir el manejo de transparencias, y complementado con técnicas que simplifican los modelos tridimensionales que tienen una enorme cantidad de polígonos a rasterizar, el *z-buffer* puede aplicarse a los polígonos que tendrán un impacto en la imagen final.

Cabe destacar que bajo su conducción por más de treinta años, Pixar realizó una serie de películas muy exitosas usando el software *RenderMan*. Este software ha sido usado en 44 de las últimas 47 películas nominadas por la Academy Award en la categoría de efectos visuales. Entre estas películas se encuentran *Avatar*, *Titanic*, *La Bella y la Bestia*, y *El Señor de los Anillos*. En sus laboratorios, fueron inventadas y publicadas una serie de tecnologías

El impacto de los algoritmos, fundamentos teóricos y software que desarrollaron Catmull y Hanrahan [...] no sólo se mide en citas o cantidad de artículos [...], sino también [...] en Premios Óscar.

fundacionales entre las cuales están composición de imágenes, *motion blur*, y simulación de ropa, entre otras.

Patrick M. Hanrahan

Pat Hanrahan es actualmente Profesor de Ciencias de la Computación e Ingeniería Eléctrica en el Computer Graphics Laboratory de la Universidad de Stanford. Obtuvo su grado de Bachelor of Science en Ingeniería Nuclear (1977) y un PhD en Biofísica (1985) de la Universidad de Wisconsin-Madison. Él fue una de las primeras personas contratadas en Pixar por Edwin Catmull. Como científico senior permaneció allí desde el año 1986 hasta el año 1989. Entre los años 1991 y 1994 fue Profesor Asociado en la Universidad de Princeton y desde el año 1994 hasta ahora está en la Universidad de Stanford.

Durante su estadía en Pixar, Hanrahan lideró el desarrollo del nuevo sistema gráfico *RenderMan*, software que permite que formas curvas puedan ser pintadas de manera realista considerando iluminación y las propiedades de los materiales (*shaders*). La idea clave fue separar el comportamiento de reflexión de la luz de la geometría del objeto y calcular el color, transparencia, y textura sobre puntos de la superficie del objeto [3]. *RenderMan* también incluyó el concepto de *z-buffering* y los algoritmos sobre los parches bicúbicos introducidos por Edwin Catmull. *RenderMan* es considerado el modelo estándar para generar efectos visuales en CGI.

La contribuciones de Hanrahan son casi innumerables; ha creado nuevos conceptos, modelos, algoritmos tanto secuenciales como paralelos, lenguajes de programación gráficos y para las GPU's, y software para *rendering* realístico de objetos, entre otras. Es difícil decidir cuales son sus aportes más importantes, pero sin duda entre estos se encuentran: (i) la creación de un nuevo método, *light field rendering,* que da al usuario la sensación de volar a través de las escenas, generando nuevas vistas desde puntos de visión arbitrarios sin información de profundidad ni geométrica, sino muestreando pedazos (*slices*) en grandes arreglos de imágenes previamente digitalizadas o pintadas [4]; (ii) técnicas para representar la piel y el pelo usando *subsurface scattering* [5]; (iii) algoritmos para modelar efectos complejos de la interacción entre distintas fuentes de luz y los objetos de la escena (iluminación global) usando *Monte Carlo ray tracing* [6]; y (iv) lenguajes para programar GPU's.

Los lenguajes para programar las GPU's (unidades de procesamiento gráfico) han sido un aporte revolucionario pues permitieron que animaciones y videojuegos tridimensionales complejos se puedan realizar en tiempo real. En este ámbito, apenas aparecieron las GPU's en los años noventa, Hanrahan y sus estudiantes extendieron el lenguaje de *shading* incluido en *RenderMan* para usar la GPU, motivando más tarde el desarrollo de versiones comerciales y el lenguaje de shading GLSL de OpenGL, la librería gráfica abierta más usada en el mundo. Más aún, en los años 2000, nuevamente junto a sus estudiantes, desarrollaron el lenguaje Brook [7], un lenguaje que permitió comenzar a usar las GPU's como poder de cálculo de propósito general y no sólo para aplicaciones gráficas. Brook motivó y condujo al desarrollo de Cuda, un lenguaje de programación de propósito general para las tarjetas gráficas NVidia.

Epílogo

El Premio Turing 2019 fue otorgado a Edwin Catmull y Patrick Hanrahan por sus contribuciones en CGI y en animación computarizada 3D. El impacto de los algoritmos, fundamentos teóricos y software que desarrollaron Catmull y Hanrahan durante sus carreras no sólo se mide en citas o en cantidad de artículos científicos destacados, sino que también se mide en Premios Óscar. Todos en nuestra vida cotidiana podemos ver ejemplos en donde sus contribuciones fueron fundamentales, por ejemplo al sentarnos al ver una película con animaciones o al jugar a nuestro videojuego favorito. Pero, sus contribuciones no sólo se limitan a la industria del entretenimiento. El desarrollo del lenguaje Brook permitió que los procesadores gráficos conocidos como GPU's, actualmente con miles de procesadores y disponibles a un precio razonable en notebooks y computadores de escritorio, pudieran ser usados como unidades de cálculo multipropósito y no sólo en el proceso de *rendering* gráfico. Es así como hoy en día se usan para para correr algoritmos incluidos en aplicaciones computacionales de alto desempeño, tales como simulaciones numéricas, análisis de imágenes en biología y medicina, entrenamiento de algoritmos de *machine learning* sobre datos masivos para aplicaciones de inteligencia artificial, entre otras. Muchos descubrimientos y avances aún por venir en el futuro se los debemos en parte al trabajo de Catmull y Hanrahan. ▪

Agradecimientos

Nuestros agradecimientos a Iván Sipirán, quien amablemente contribuyó con las imágenes para las Figuras 2 y 3 de este artículo.

REFERENCIAS

[1] Andrew Utterson. A Computer Generated Hand. Ensayo para el National Film Registry. https://www.loc.gov/static/programs/national-film-preservation-board/documents/computer_hand2.pdf (último acceso: 14 de abril de 2021).

[2] E. Catmull, AR Smith. 3-D transformations of images in scanline order. ACM SIGGRAPH Computer Graphics 14 (3):279-285. 1980.

[3] RA Drebin, L Carpenter, P Hanrahan. Volume rendering. ACM SIGGRAPH Computer Graphics, 22 (4):65-74. 1988.

[4] Marc Levoy, Pat Hanrahan. Light field rendering. Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, pages 31-42. 1996.

[5] SR Marschner, HW Jensen, M Cammarano, S Worley, P Hanrahan. Light scattering from human hair fibers. ACM Transactions on Graphics (TOG) 22 (3):780-791. 2003.

[6] State of the Art in Monte Carlo Ray Tracing for Realistic Image Synthesis. SIGGRAPH 2001 Course 29. 2001. Available from: https://www.researchgate.net/publication/2872516_State_of_the_Art_in_Monte_Carlo_Ray_Tracing_for_Realistic_Image_Synthesis#fullTextFileContent (último acceso: 20 de mayo de 2021).

[7] Ian Buck, Tim Foley, Daniel Horn, Jeremy Sugerman, Kayvon Fatahalian, Mike Houston, Pat Hanrahan. Brook for GPUs: stream computing on graphics hardware. ACM Transactions on Graphics (TOG) 23 (3):777-786. 2004.

Historia y evolución de la inteligencia artificial

ANDRÉS ABELIUK

Profesor Asistente del Departamento de Ciencias de la Computación de la Universidad de Chile. Ph.D en Ciencias de la Computación por la Universidad de Melbourne, Australia. Líneas de investigación: computación social e inteligencia colectiva, análisis de redes sociales e impacto de la inteligencia artificial en la sociedad.

aabeliuk@dcc.uchile.cl

CLAUDIO GUTIÉRREZ

Profesor Titular del Departamento de Ciencias de la Computación de la Universidad de Chile. Investigador Senior del Instituto Milenio Fundamentos de los Datos. Licenciado en Matemáticas por la Universidad de Chile y Ph.D. en Computer Science por la Wesleyan University. Líneas de investigación: fundamentos de los datos, bases de datos, lógica aplicada a la computación y semántica de la Web.

cgutierr@dcc.uchile.cl

El primer programa de IA

En 1842, la matemática y pionera de la informática, Ada Lovelace, programó el primer algoritmo destinado a ser procesado por una máquina. Adelantada a su época, Ada especuló que la máquina "podría actuar sobre otras cosas además de los números... el motor (la máquina) podría componer piezas musicales elaboradas y científicas de cualquier grado de complejidad o extensión". Décadas más tarde, la visión de Ada es una realidad gracias a la Inteligencia Artificial (IA). Sin embargo, un hito considerado como el momento fundacional de la "inteligencia artificial", tanto del término como del campo de estudio, es una conferencia en Darmouth el año 1956 organizada por John McCarthy, Marvin Minsky, Claude Shannon y Nathaniel Rochester [1]. En ella, los organizadores invitaron a unos diez investigadores para formalizar el concepto de inteligencia artificial como un nuevo campo de estudio científico. Pioneros de la IA, cuatro de los asistentes fueron posteriormente galardonados con el premio Turing (a menudo denominado Premio Nobel de informática) por sus contribuciones a la IA. Una idea común entre los asistentes, y profundamente arraigada hasta el día de hoy en el estudio de la IA, es que el pensamiento es una forma de computación no exclusiva de los seres humanos o seres biológicos. Más aún, existe la hipótesis de que la inteligencia humana es posible de replicar o simular en máquinas digitales.

Ese mismo año dos de los participantes de la conferencia, Alan Newell y Herbert Simon, publican lo que es considerado el primer programa computacional de inteligencia artificial [2]. El programa "Logic Theory Machine" es capaz de descubrir demostraciones de teoremas en lógica simbólica. La idea principal es que a través de la combinación de simples operaciones primitivas, el programa puede ir construyendo expresiones cada vez más complejas. El desafío computacional radica en encontrar la combinación de operaciones que demuestran un teorema dado, entre una cantidad exponencial de posibles combinaciones. La contribución de los autores fue idear un enfoque heurístico, o de reglas generales, que permiten recortar el árbol de búsqueda de manera "inteligente" y encontrar una solución en la mayoría de los casos, pero no siempre. La introducción de los procesos heurísticos han influenciado enormemente la ciencia de la computación y según los mismos autores, son la magia central en toda resolución de problemas humanos. No es coincidencia que esta tesis provenga de Herbert Simon, quien recibió el Nobel en economía por la provocadora idea de modelar el comportamiento humano, no como un agente "homo economicus" totalmente racional, sino que con "racionalidad limitada" cuya toma de decisiones es principalmente heurística [3].

Dos paradigmas de investigación en IA

IA simbólica

La búsqueda heurística fue un pilar clave para los avances de la IA en sus comienzos. Todo tipo de tareas de resolución de problemas, como probar teoremas y jugar ajedrez, implican tomar decisiones que se pueden modelar como un árbol de decisiones que debe ser recorrido para encontrar una estrategia que resuelva el problema. Los algoritmos de búsqueda heurística son parte de una colección de métodos que se basan en representar el conocimiento implícito o procedimental que poseen los humanos de forma explícita, utilizando símbolos y reglas (legibles por humanos) en programas informáticos. La "IA simbólica" demostró ser muy exitosa en las primeras décadas de la IA logrando codificar

en “sistemas expertos” el razonamiento humano en dominios de conocimiento específico. Un ejemplo son los sistemas de apoyo de diagnóstico médico a través de motores de inferencia y bases de conocimientos que resumen el conocimiento médico basado en evidencia. Uno de los logros más populares de la IA simbólica culmina con la derrota del campeón mundial de ajedrez en 1997, Garry Kasparov, por el computador Deep Blue de IBM [4] (ver infografía de línea de tiempo en la Figura 1).

IA conexionista

Paralelamente con la emergencia de la IA simbólica, que modela la mente humana como si fuese una computadora procesadora de símbolos, existe otra escuela de pensamiento que se basa en modelar la biología del cerebro que está compuesto por redes neuronales biológicas. Frank Rosenblatt (psicólogo) en 1958 propuso el *perceptrón*, una generalización de la neurona McCulloch-Pitts que podía “aprender” a través de coeficientes de ponderación para cada entrada de la neurona. Hasta el día de hoy, el perceptrón es la unidad fundamental para muchas de las redes neuronales artificiales e impulsa el paradigma conocido como IA conexionista. A pesar de su promesa, la investigación en redes neuronales se detuvo por falta de financiamiento y una sobreexpectación no cumplida. Hechos que parcialmente son atribuidos a una malinterpretada exposición de las limitaciones y fortalezas del perceptrón en un libro por pioneros de la IA simbólica, Marvin Minsky y Seymour Papert en 1969 [5]. No fue hasta comienzos de 1980, que Geoffrey Hinton (Premio Turing en 2018) y colegas redescubren y popularizan el método llamado *retropropagación* [6]; el algoritmo central detrás de la *búsqueda heurística* (estilo IA simbólica) que logra encontrar los parámetros del modelo que minimizan su error, así permitiendo que una red neuronal de múltiples capas aprenda a partir de datos. Este avance resuelve las limitaciones de los perceptrones de Rosenblatt y crea un resurgimiento en la investigación del aprendizaje profundo (ver Figura 1).

1842

De los números a la poesía

La matemática Ada Lovelace fue la primera en ver el potencial de las computadoras más allá de las matemáticas.

1921

Se introduce la palabra “Robot”

Karel Čapek, un dramaturgo checo, lanzó su obra de ciencia ficción “Rossum’s Universal Robots”, donde exploró el concepto de personas artificiales a las que llamó robots, que proviene de la palabra “robota” (esclavo).

1943

Las neuronas se vuelven artificiales

El primer modelo matemático de la neurona fue propuesto por Warren McCulloch y Walter Pitts.

1950

Turing Test

Alan Turing propone un test para saber si una máquina exhibe un comportamiento inteligente.

1956

La “IA” nace

El término “inteligencia artificial” es acuñado en una conferencia en la Universidad de Dartmouth organizada por John McCarthy.

1956

El primer programa de IA

Allen Newell, Herbert Simon y Cliff Shaw fueron coautores de Logic Theorist, el primer programa informático de inteligencia artificial.

1961

Unimate

El robot industrial, Unimate, inventado por George Devol, se convirtió en el primero en trabajar en una línea de montaje de General Motors.

1964

Eliza

Joseph Weizenbaum, científico informático de MIT, desarrolló ELIZA, el primer chatbot que podía conversar funcionalmente en inglés con una persona.

1969

El problema del XOR

Marvin Minsky y Seymour Papert exploran en un libro las fortalezas y limitaciones de los perceptrones, la más importante siendo la incapacidad de implementar la función lógica XOR.

1974 - 1980

Invierno IA

Muchos comienzos en falso y callejones sin salida dejan a la investigación en IA sin fondos y con poco interés.

1986

Aprendiendo a aprender con retropropagación

En un artículo muy influyente, Rumelhart, Hinton, y Williams, popularizan el algoritmo de retropropagación para entrenar redes neuronales multicapa.

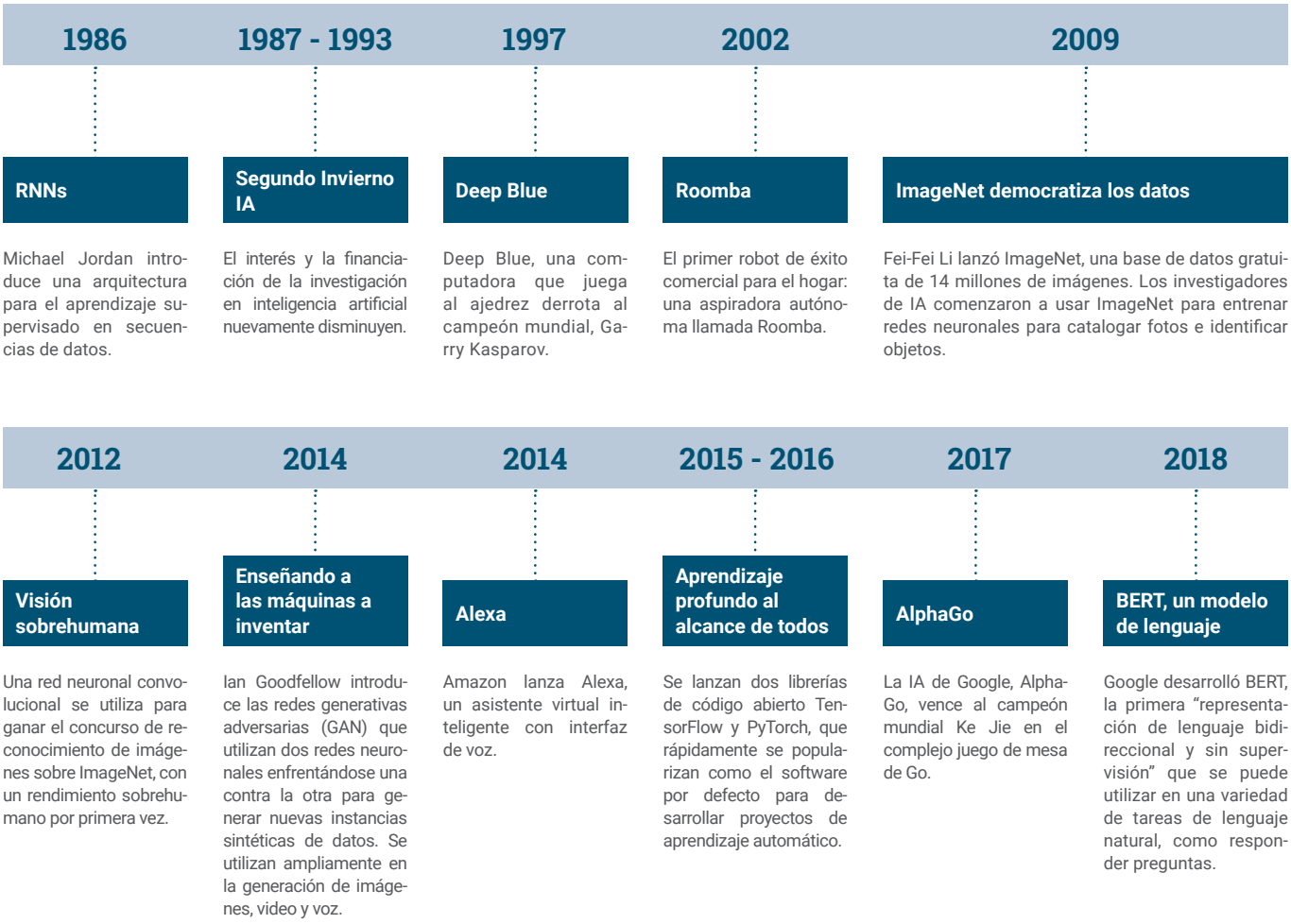


Figura 1. Historia de la inteligencia artificial.

La revolución del aprendizaje profundo

En la década del 2010, dos cosas harían posible la revolución de aplicaciones de redes neuronales y algoritmos de aprendizaje profundo. Primero, los avances de hardware especializado han acelerado drásticamente el entrenamiento y el rendimiento de las redes neuronales y reducido su consumo de energía. Segundo, el aumento de datos abiertos disponibles online y servicios de bajo costo para etiquetar datos vía *crowdsourcing* impulsan el desarrollo de la IA. La Figura 2 muestra cómo los conceptos de IA, aprendizaje automático (*machine learning*) y aprendizaje profundo (*deep learning*) se relacionan el uno con el otro.

Como consecuencia de estos avances, se desarrollaron aplicaciones basadas en las redes neuronales donde la IA simbólica no tuvo éxito. Por ejemplo en aplicaciones de visión, como reconocimiento facial y detección de cáncer, y en aplicaciones de lenguaje, como la traducción de idiomas y asistentes virtuales. En 2015, Microsoft Research utiliza una arquitectura de red neuronal para categorizar imágenes con una mayor precisión que el humano promedio [7]. Al siguiente año, el sistema AlphaGo de DeepMind se corona maestro de Go tras vencer al campeón mundial, Lee Sedol [8]. Este suceso es impactante ya que en el Go hay en promedio alrededor de 300 movimientos posibles que se pueden hacer en cada turno, mientras que en el ajedrez es cercano a 30 movimientos. En otras palabras, el árbol de búsqueda del Go tiene un factor de ramificación de un orden de magnitud mayor al ajedrez, razón principal por la cual la IA simbólica, por sí sola, falló en desarrollar un programa para jugar Go.

Limitaciones de la IA

Un aspecto clave y poderoso de las redes neuronales es que no requieren que se especifiquen las reglas del dominio a modelar; las reglas se aprenden a partir de los datos de entrenamiento. La falta de conocimiento de alto nivel embebido en el sistema por expertos humanos, como es el caso de la IA simbólica, se contrarresta con la capacidad de inferir estadísticamente un modelo del dominio a partir de suficientes datos. Sin embargo, una desventaja importante de las redes neuronales es que requieren grandes recursos computacionales y cantidades enormes de datos. Por ejemplo, se estima que replicar los experimentos de AlphaGo costaría alrededor de 35 millones de dólares sólo en poder computacional [9]. Por otro lado, los datos deben ser cuidadosamente "curados" para ser representativos y así poder generalizar correctamente y no producir resultados sesgados, como ha sido el caso en textos sexistas y racistas generados a partir de modelos de lenguaje [10]. Por otro lado, mientras que programas de software basados en reglas explícitas son fáciles de rastrear y comprender cómo llegaron a tomar ciertas decisiones, no se puede decir lo mismo de los algoritmos de aprendizaje profundo que debido a su alta complejidad son difíciles de interpretar y comunicar por humanos. Estas limitaciones son uno de los grandes desafíos en la IA y hay mucha investigación activa en estas direcciones [11,12].

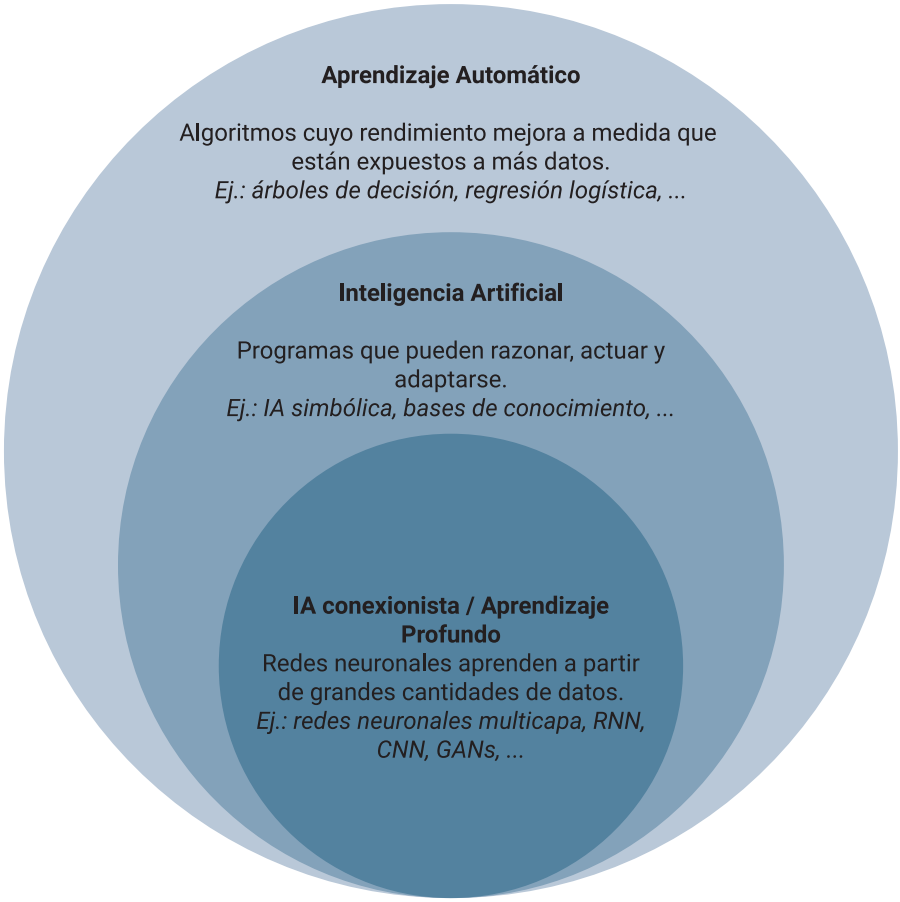


Figura 2. Diagrama de Venn que muestra la relación entre distintas subáreas de la inteligencia artificial.

Democratizando la IA

Desde que el aprendizaje profundo recuperó prominencia alrededor del 2010, los softwares gratuitos y de código abierto especializados para el aprendizaje profundo han sido enormemente responsables de impulsar el campo hacia adelante. Desde las primeras librerías creadas por equipos académicos, Caffe y Theano, hasta las actuales dominantes, PyTorch y TensorFlow, respaldadas por Facebook y Google, respectivamente, el acceso a estos softwares de código abierto han facilitado el cambio hacia la innovación tecnológica impulsada por el aprendizaje automático. Tanto en la investigación de vanguardia como en la creación de aplicaciones por la industria, la democratización de la IA reduce las barreras de entrada para que las personas y organizaciones puedan ingresar al apasionante mundo de la IA con poca o nada de inversión financiera. Pueden aprovechar los datos y algoritmos disponibles públicamente para comenzar a experimentar la construcción de modelos de IA y a la vez contribuir a expandir las bases de datos públicas y poner a disposición nuevas soluciones.

Como ejemplo del poder de democratizar datos, en el 2009 el proyecto ImageNet, liderado por la investigadora Fei-Fei Li, puso a disposición del público una gran base de datos visual que ayudó a investigadores a crear modelos más rápidos y precisos de reconocimiento visual de objetos. Esta colección de imágenes se convirtió rápidamente en una competencia anual (ahora organizada en Kaggle) para ver qué algoritmos podían identificar objetos en las imágenes

La búsqueda heurística fue un pilar clave para los avances de la IA en sus comienzos.

con la tasa de error más baja. El 2012, el primer equipo en usar redes neuronales en la competencia venció el estado del arte con una precisión récord. La arquitectura propuesta por integrantes del laboratorio de Geoffrey Hinton en la Universidad de Toronto, Red Neuronal Convolucional [13], fue inspirada por las características estructurales y fisiológicas de la visión animal. Hoy en día, estas redes neuronales están en todas partes: se usan para etiquetar las fotos en plataformas sociales; los vehículos autónomos las utilizan para detectar objetos; y se usan para digitalizar textos.

Desde entonces, se han introducido una multitud de nuevos conjuntos de datos estimulando investigación en subcampos de la IA como el procesamiento de lenguaje natural (NLP) y reconocimiento de voz y audio. La arquitectura precursora en NLP, es la Red Neuronal Recurrente, que usa datos secuenciales y se distingue por su "memoria", ya que al iterar sobre la entrada, mantiene un estado interno que codifica información sobre los elementos anteriores dentro de la secuencia e influenciando el output actual [14]. El procesamiento de lenguaje natural juega un papel vital en muchos sistemas, desde el análisis de curriculums para la contratación, hasta asistentes virtuales y detección de spam. Sin embargo, el desarrollo y la implementación de la tecnología de NLP no es tan equitativo como parece. Aunque se hablan más de 7000 idiomas en todo el mundo, la gran mayoría de los avances tecnológicos son aplicados al inglés. Una iniciativa para contrarrestar esta inequidad es liderada por Jorge Pérez, académico del Departamento de Ciencias de la Computación (DCC) de la Universidad de Chile, que junto a estudiantes han puesto a libre disposición de la comunidad un modelo de lenguaje en español [15].

IA en Chile

Describiremos a grandes rasgos el desarrollo actual de la IA en Chile en tres áreas: empresarial, investigación académica, y enseñanza y propuestas.

A nivel empresarial las técnicas de IA ya están comenzando a ser un *commodity*, esto es, están a disposición en el mercado regular y se están usando de manera generalizada (particularmente en lo que respecta a aprendizaje por medio de datos). Otra pregunta es si hay desarrollos "novedosos". A manera de ejemplo nombraremos cuatro. NotCo, cuyo logro es "combinar la inteligencia artificial con el conocimiento del mundo vegetal para crear productos". Usan técnicas de análisis de datos y visualización innovadoras, pero hay poca información sobre su nivel innovador de IA/ML. El otro ejemplo es Fintual. Aquí se usan bastantes cosas que se pueden considerar IA, por ejemplo, "bots de inversión" que permiten seguir ciertos índices para invertir de

Un aspecto clave [...] de las redes neuronales es que no requieren que se especifiquen las reglas del dominio a modelar; las reglas se aprenden a partir de los datos de entrenamiento.

forma pasiva y a bajo costo. El tercer ejemplo es CornerShop que usa tecnologías de datos y analítica para su diseño de operaciones. Finalmente, un ejemplo de una empresa más pequeña es Zippedi, orientada a robótica de almacenes para optimizar digitalmente las estanterías. Hay también muchas otras empresas tipo *startup* que están haciendo cosas tipo *chatbots*, aplicaciones de procesamiento de imágenes, bioinformática, etc., la mayoría aplicando investigación ya consolidada (no desarrollando).

Respecto de la investigación dedicada a la IA propiamente tal (esto es, publicando regularmente en revistas o conferencias de IA) son pocos los grupos a lo largo del país. Destacamos IALab de la Pontificia Universidad Católica de Chile, que tiene varios años y buena infraestructura (particularmente su *cluster* de GPUs para IA). Su fuerte es visión computacional y robótica. Otro grupo es el de Inteligencia Computacional del Departamento de Ingeniería Eléctrica (UChile) que está centrado en robótica y visión, y procesamiento de señales y aprendizaje en este campo. En el DCC (UChile) hay un grupo (ReLeLa) centrado en IA y NLP. La Universidad de Concepción recientemente creó un grupo de IA enfocado a Sistemas Multiagente y Robótica. Por otra parte, hay muchos grupos en diversas universidades dedicados más bien a aplicaciones de AI en diferentes áreas, como empresarial, comercial, científica, social, etc. y luego publican en esas disciplinas. Por ejemplo, el Instituto Data Science de la Universidad del Desarrollo (UDD) aplica técnicas de IA en proyectos asociados a la minería y agricultura. Finalmente, hay muchxs investigadores que trabajan más bien solos o con colegas de otras instituciones en diversas universidades a lo largo del país.

Respecto de la enseñanza, han proliferado los cursos de IA, así como diplomados y magíster en el área dictados por diferentes universidades y organizaciones. Esto muestra que la IA se ha convertido en un *boom* en Chile, con los claros y oscuros propios de un *boom*. En este marco diferentes organizaciones e instituciones discuten sobre los usos de IA en diferentes áreas, entre ellos, la Comisión Desafíos del Futuro del Senado, las universidades, las Fuerzas Armadas, el Ministerio de Ciencia y Tecnología, etc. Se han elaborado diversos documentos. Algunos ejemplos son: "Inteligencia Artificial para Chile. La urgencia de desarrollar una estrategia", del Senado de la República; "Ejército Virtual" de la Academia Politécnica Militar, y "Política Nacional de Inteligencia Artificial" del Ministerio de Ciencia.

Ética, alcances y limitaciones de la IA

Como toda tecnología, la IA trae aparejada dilemas éticos. En el caso de la IA esto se agranda por el poder transformador de la realidad que puede traer aparejado esta tecnología. Así es que hoy, al igual que la investigación biomédica desde siempre, el test de la ética debe ser aplicado a los desarrollos de IA. Esto se refiere particularmente a funcionalidades donde existen máquinas y aparatos con "inteligencia" o habilidades de simulación de lo humano que sobrepasan con creces las de los humanos. Y las preguntas fundamentales van por el lado del marco ético para los desarrollos en esta disciplina. Mencionaremos algunos de los principales:

1. *La IA y la economía* [16]. Aquí aparecen temas como los usos de la IA en el mundo del trabajo: por ejemplo, ¿dónde están los límites de los flujos de trabajo automatizado donde hay personas involucradas? Y la pregunta fundamental del área: ¿cómo distribuiremos la riqueza creada por las máquinas?

2. *La IA y la sociedad* [17]. ¿Cómo afectan las máquinas inteligentes la relación entre los seres humanos? ¿Quiénes decidirán los usos de las máquinas inteligentes? ¿Quién y cómo controlar los sesgos (introducidos intencional o no intencionalmente) a las máquinas? ¿Cuáles son los límites (o no existen) al desarrollo de ese tipo de proyectos?

3. *La IA y los humanos*. ¿Cómo afectarán las máquinas inteligentes nuestro comportamiento? ¿Hasta qué nivel es permisible "ensamblar" esas máquinas con nuestra biología?

4. *La IA y el medio ambiente*. ¿Cuáles son los límites razonables de uso de recursos para estos proyectos?

5. *Seguridad, usos militares* [18]. ¿Qué es necesario y cómo regular este ámbito, tradicionalmente complejo de regular?

6. *Superinteligencia* [19]. ¿Qué derechos y deberes tendrán estos robots? ¿Quién es responsable por sus desarrollos y usos? ¿Qué nivel de decisiones se les permitirá tomar en asuntos humanos?

Hay miles de otras preguntas. Uno podría replicar todas las de la ética clásica, pues en definitiva lo que está ocurriendo con la IA débil al menos, es la realización de gran parte de los proyectos clásicos de simulación de facetas de lo humano. ▪

REFERENCIAS

[1] James Moor. "The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years". *AI Magazine* 27(4), 2006. https://doi.org/10.1609/aimag.v27i4.1911.

[2] A. Newell y H. Simon. "The Logic Theory Machine – A Complex Information Processing System". *IRE Transactions on Information Theory* 2, 1956.

[3] Wheeler, Gregory. "Bounded Rationality". *The Stanford Encyclopedia of Philosophy*, 2020. https://plato.stanford.edu/archives/fall2020/entries/bounded-rationality/.

[4] Hansen Hsu. AI and Play, part 1: How Games Have Driven two Schools of AI Research, Computer History Museum, 2020. https://computerhistory.org/blog/ai-and-play-part-1-how-games-have-driven-two-schools-of-ai-research/.

[5] Minsky, Marvin, y Seymour A. Papert. *Perceptrons: An Introduction to Computational Geometry*. MIT press, 2017.

[6] Rumelhart, D. E., Hinton, G. E., y Williams, R. J. Learning Representations by Back-Propagating Errors. *Nature,* 1986.

[7] He, Kaiming, et al. "Delving Deep into Rectifiers: Surpassing Human-Level Performance on Imagenet Classification". *Proceedings of the IEEE international conference on computer vision*, 2015.

[8] Silver, D., Huang, A., Maddison, C. et al. Mastering the Game of Go with Deep Neural Networks and Tree Search. Nature 529, 2016.

[9] DeepMind's Losses and the Future of Artificial Intelligence. WIRED, 2019. https://www.wired.com/story/deepminds-losses-future-artificial-intelligence/.

[10] Zou, J. y Schiebinger, L. AI can be Sexist and Racist – It's Time to Make it Fair. Nature 559, 324–326, 2018.

[11] Xie, Ning, et al. "Explainable Deep Learning: A Field Guide for the Uninitiated". *arXiv preprint,* 2020.

[12] Mehrabi, Ninareh, et al. "A Survey on Bias and Fairness in Machine Learning". *arXiv preprint,* 2019.

[13] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet Classification with Deep Convolutional Neural Networks". *Advances in neural information processing systems* 25, 2012.

[14] Mikolov, Tomáš, et al. "Recurrent Neural Network based Language Model". Eleventh annual conference of the international speech communication association, 2010.

[15] Cañete, José, Gabriel Chaperon, Rodrigo Fuentes, y Jorge Pérez. "Spanish pre-trained Bert Model and Evaluation Data". *PML4DC at ICLR* 2020, 2020.

[16] Egana-delSol, Pablo. "The Future of Work in Developing Economies: What can we learn from the South?". *Available at SSRN 3497197,* 2019.

[17] Tomašev, N., Cornebise, J., Hutter, F. et al. AI for Social Good: Unlocking the Opportunity for Positive Impact. Nat Commun 11, 2468, 2020.

[18] Toby Walsh A.I. Expert, Is Racing to Stop the Killer Robots. The New York Times, 2019. https://www.nytimes.com/2019/07/30/science/autonomous-weapons-artificial-intelligence.html.

[19] Alfonseca, M., Cebrián, M., Anta, A. F., Coviello, L., Abeliuk, A., y Rahwan, I. Superintelligence cannot be Contained: Lessons from Computability Theory. *Journal of Artificial Intelligence Research*, 2021.

Hacia una política chilena de inteligencia artificial, nacida en contexto de pandemia

ANDREA RODRÍGUEZ

Integrante de la Comisión Asesora para la Política Nacional de Inteligencia Artificial. Vicerrectora de Investigación y Profesora Titular del Departamento de Ingeniería Informática y Ciencias de la Computación de la Universidad de Concepción. Investigadora Asociada del Instituto Milenio Fundamentos de los Datos. PhD. in Spatial Information Science and Engineering por la Universidad de Maine, Estados Unidos.

andrea@udec.cl

Motivado por el impacto actual y proyectado a nivel mundial, tanto en lo económico y social de la inteligencia artificial, junto con el diagnóstico entregado por la Comisión Desafíos del Futuro del Senado que levantó la necesidad de una Estrategia Nacional de Inteligencia Artificial, el Gobierno de Chile encarga al Ministerio de Ciencia, Tecnología, Conocimiento e Innovación (CTCI), a mediados del año 2019, la elaboración de una Política Nacional de Inteligencia Artificial y su Plan de Acción, proceso que al momento de escribir este artículo no está concluido. Este artículo describe los lineamientos generales de la política presentada a consulta pública, la cual se encuentra en su etapa final de elaboración. Cabe señalar que la situación de pandemia no sólo atrasó el proceso, sino que además creó un escenario donde será necesaria la decisión y convicción para impulsar acciones que demandarán, inevitablemente, recursos económicos.

La elaboración de una política de inteligencia artificial requiere hacer explícitas definiciones que permitan comunicar con claridad la visión de lo que se espera alcanzar. Algo comúnmente aceptado es definir inteligencia artificial como una disciplina que aborda la creación de métodos computacionales que realizan tareas consideradas inteligentes, en específico, que razonan, se adaptan y actúan. Partiendo de esto, uno puede entender qué o qué no abarca inteligencia artificial. Inteligencia artificial no es equivalente a transformación digital, automatización, sensorización (*Internet of Things*) o robótica. Aunque relacionados, mezclar temas nos llevan a confundir el real avance que podemos tener en inteligencia artificial. Por ejemplo, la transformación digital en distintos ámbitos puede llevarse a cabo sin haber logrado avances importantes en inteligencia artificial.

Contexto internacional respecto a políticas de inteligencia artificial

En el contexto internacional, varios países han elaborado sus propias estrategias para el fomento de la inteligencia artificial. Tal es así, que el AI Index Report 2021 de la Universidad de Stanford da cuenta de 32 países que ya han elaborado y 22 países que están en proceso de elaborar estrategias de inteligencia artificial. En un comienzo Canadá, China, Japón, entre otros, en el año 2017 establecieron estrategias y objetivos de fomento de la inteligencia artificial. Canadá enfatiza el aporte desde la academia, financiando institutos, investigadores, investigación en inteligencia artificial y sociedad, y un programa nacional de encuentros. China se plantea como meta el liderazgo a partir de la superación de brechas y de potenciar el desarrollo de tecnología y fomento de innovación basado en inteligencia artificial en el sector privado. Japón establece etapas de desarrollo, desde la utilización de datos e inteligencia artificial en la industria de servicios relacionados, para pasar a su uso público y expansión, y terminar con la creación de un ecosistema que potencie la integración.

En el año 2018 se suman Francia, Alemania, Reino Unido con sus propias estrategias de inteligencia artificial. Francia resalta los aspectos éticos y de inclusión, valorando el desarrollo económico basado en datos, proponiendo la creación de un número específico de centros interdisciplinarios y definiendo como sectores estratégicos salud, medio ambiente, transporte/movilidad, defensa y seguridad. Alemania por su lado, enfatiza la necesidad de un desarrollo de inteligencia artificial que considere a la sociedad y el desarrollo sostenible, incorporando la necesidad de medidas de monitoreo y diagnóstico de las aplicaciones y la penetración de la inteligencia artificial en la sociedad. Reino Unido marca una diferencia entre las acciones del Gobierno y la industria, donde la inteligencia artificial pasa a ser uno de los grandes desafíos de su política industrial, promoviendo la innovación y productividad en los distintos sectores junto a la generación de talento. A nivel latinoamericano, y en ese mismo año, México es el primer país en elaborar una estrategia con el objetivo de impulsar su liderazgo en la materia. La estrategia mexicana propone el desarrollo de un marco de gobernanza multisectorial, un levantamiento de usos, necesidades industriales y mejores prácticas en el Gobierno, además de incorporar el trabajo con expertos que permitan la continuidad de las iniciativas.

Un año más tarde, en 2019, Rusia y Estados Unidos, entre otros países, presentan sus estrategias. Rusia pone

énfasis en intereses nacionales, en una proyección hasta el año 2030. Esto incluye iniciativa tecnológica nacional, proyectos departamentales y programas como el Economía Digital de la Federación de Rusia. Estados Unidos por otra parte, prioriza la necesidad de que el gobierno federal invierta en investigación y desarrollo en inteligencia artificial y garantizar normas técnicas para el desarrollo seguro y despliegue de tecnologías. También el año 2019 Uruguay lanzó a consulta pública una estrategia de inteligencia artificial que identifica como pilares la gobernanza de la política, el desarrollo de capacidades de inteligencia artificial, inteligencia artificial y ciudadanía, y el uso responsable. Interesante es hacer notar que Uruguay es reconocido por el Government AI Readiness Index del 2020 desarrollado por Oxford Insights y el International Research Development Center como el país mejor posicionado en América Latina en el uso responsable de inteligencia artificial por parte del Gobierno, seguido en la región en este ranking por Chile.

Terminando esta acotada revisión, en el año 2020 España presenta una estrategia, recomendando la coordinación entre instrumentos de fomento y agencias estatales de I+D+i en sectores estratégicos para la economía y sociedad, tales como educación, ciudad y territorio, salud, energía, seguridad y turismo e industrias creativas.

Existiendo diversidad en la forma en que se gestaron las diversas estrategias nacionales, aspectos comunes a varias de ellas incluyen la necesidad de contar con profesionales del área, el fomento de las capacidades en infraestructura y disponibilidad de datos, el apoyo a la investigación y la innovación, la definición de áreas estratégicas de aplicación, y la consideración de aspectos éticos y de impacto en la sociedad. Algunos de estos países, en particular países con mayores recursos, explicitaron en su momento los recursos necesarios, los que superan por mucho los cientos de millones de dólares en periodos que van desde 4 a 13 años. Todo lo anterior en condiciones prepandemia. Habrá que esperar para dimensionar cómo estas estrategias y sus plazos puedan verse afectados en la situación actual.

Inteligencia artificial no es equivalente a transformación digital, automatización, sensorización (Internet of Things) o robótica.

Eje	Temas asociados
Factores habilitantes	• Capital Humano • Infraestructura • Datos
Desarrollo y adopción	• Generador de indicación y uso y adopción en sector público y privado • Valorización en productividad científica • Vinculación al sector privado • Fomento a la innovación y emprendimiento • Consideración del medio ambiente
Ética, aspectos normativos y efectos sociales y económicos	• Uso seguro y respetuoso de las personas • Ciberseguridad • Monitoreo del efecto en el empleo • Protección de datos • Propiedad intelectual

Figura 1. Ejes fundamentales de la política chilena de inteligencia artificial.

Lineamientos de la política nacional de inteligencia artificial

A nivel nacional, y sin entrar en mayor detalle del proceso, la elaboración de la política chilena incluyó diferentes actores. Por un lado, se creó un comité interministerial y se organizó el trabajo de coordinación y escritura del trabajo en un grupo gestionado por el Ministerio de CTCI. El comité asesor de expertos jugó el rol de asesor entregando ideas y revisando la propuesta de política, y donde participan algunos miembros que fueron anteriormente convocados por la Comisión Desafíos del Futuro del Senado. Por otro lado, se realizó un proceso participativo durante el 2020 a través de mesas de trabajo, charlas y una consulta ciudadana amplia en base al borrador de la política.

En su versión preliminar y llevada a consulta pública, la política chilena de inteligencia artificial se proyecta hasta el año 2030 y establece como misión el *"empoderar al país en el uso y desarrollo de sistemas de inteligencia artificial, propiciando el debate sobre sus dilemas éticos y sus consecuencias regulatorias, sociales y económicas"*. Usa como principios transversales el desarrollo de inteligencia centrado en las personas, fomentando el desarrollo sostenible, enfatizando los aspectos de seguridad e inclusión, e insertada globalmente.

La política chilena de inteligencia artificial [...] usa como principios transversales el desarrollo de inteligencia centrado en las personas, fomentando el desarrollo sostenible, enfatizando los aspectos de seguridad e inclusión, e insertada globalmente.

Con un enfoque que extrae aspectos comunes a las políticas o estrategias de otros países, la política chilena de inteligencia artificial fue diseñada en torno a tres ejes principales (ver Figura 1). El eje de factores habilitantes considera los factores o elementos necesarios para el desarrollo de la inteligencia artificial en el país, abarcando principalmente capital humano, datos e infraestructura. El eje de desarrollo y adopción considera las formas de promoción del uso de la tecnología, la adopción en sectores relevantes para el país, y los roles que cumple la investigación; además la transferencia, emprendimiento e innovación, tanto en el sector público como privado. El eje de ética, aspectos normativos y efectos sociales y económicos, por su parte, plantea la discusión en torno al efecto de la inteligencia artificial en el campo laboral y la discusión de los requisitos que den garantía de su uso seguro y responsable socialmente.

En la formación de capital humano del eje de factores habilitantes, la política establece como objetivos la formación a distintos niveles, desde establecer competencias computacionales a nivel escolar, promover e incorporar la inteligencia artificial como una disciplina transversal a nivel técnico y profesional, y apoyar postgrados en el área, todo esto con un enfoque de formación continua y de capacitación para la conversación laboral que resulte necesaria.

El objetivo a nivel de profesionales y expertos es alcanzar los niveles de la OCDE, teniendo en cuenta que se estima que en Chile existe una persona dedicada a investigación y desarrollo por cada mil personas de la fuerza laboral; mientras que el promedio de la OCDE es ocho. En este mismo eje, se plantea esencialmente disponer de conectividad de calidad desplegada a nivel territorial, generando una hoja de ruta de infraestructura de almacenamiento y cómputo para el desarrollo científico-tecnológico. Finalmente, se plantea la relevancia de la disposición de datos científicos, y datos del sector público y privado que fomenten el desarrollo y el valor agregado de las herramientas basadas en aprendizaje.

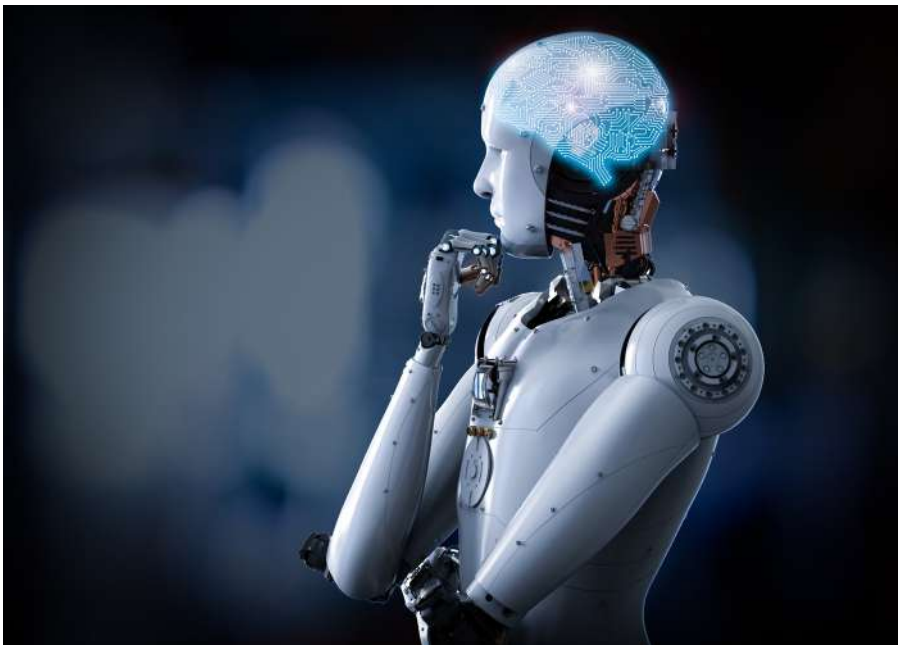
En el eje de desarrollo y adopción, se plantea en forma transversal generar indicadores que valoren adecuadamente la investigación en el área e indicadores que permitan monitorear la adopción de tecnología en el sector público y privado. Para la transferencia tecnológica, la innovación y el emprendimiento, se reconoce la necesidad de fomentar la relación academia-industria, la generación armónica de una comunidad de emprendedores y el fomento al emprendimiento con base científico-tecnológico en inteligencia artificial. Con el objetivo de lograr un mejoramiento de servicios públicos, se plantea generar un plan de ruta partiendo por el sistema de compras públicas, y en el sector privado fomentar la capacitación e inserción de capacidades en el sector.

Alineado con el principio de sustentabilidad, la política propone fomentar la investigación, el desarrollo y el uso de sistemas de IA con consideraciones de eficiencia e impacto en el medio ambiente. Así mismo, establece el potencial desarrollo de aplicaciones de inteligencia artificial asociado al monitoreo del medio ambiente en consonancia con la existencia de datos impulsado por el observatorio de cambio climático.

En el tercer eje de ética, aspectos normativos y efectos sociales y económicos, la política enfatiza el desarrollo y uso de inteligencia artificial que sea concordante con los derechos fundamentales, no discriminatorio, inclusivo y respetuoso de las normas de protección de datos personales. Así mismo, plantea el monitoreo del empleo y proveer mecanismos para resolución de conflictos con los trabajadores. El desarrollo seguro de la inteligencia artificial se asocia a posicionar la ciberseguridad como un componente central de los sistemas de inteligencia artificial y donde a su vez la inteligencia artificial puede aportar. Respecto a temas regulatorios, se releva la importancia de la inserción en la discusión regulatoria a nivel internacional, asociando además los temas de propiedad intelectual con el impulso económico que el desarrollo de inteligencia artificial puede lograr.

Comentarios generales

La política chilena estructurada en los tres ejes propuestos debiera cubrir brechas y posicionarnos a un nivel de liderazgo, al menos a nivel de América Latina. Pero estas brechas parecen no



[La inteligencia artificial] abre la posibilidad de que aumenten las brechas e inequidades entre aquellos que tengan o no tengan el poder de esta tecnología. Enfrentar estos desafíos requiere [...] verla como una tecnología que ha sido creada por la inteligencia humana y que debe estar al servicio de la sociedad.

estar totalmente claras aún, y varias de las iniciativas apuntan a generar indicadores que permitan hacer seguimiento de la adopción y el impacto de la inteligencia artificial en el país. Chile, al igual que otros países de la región, tiene un nivel de digitalización heterogéneo, con un claro mayor desarrollo en tecnologías de información y de comunicación en torno a las grandes urbes. Si en algo la pandemia nos ha hecho avanzar, es en forzar una mayor cobertura para la conectividad, y hacer evidente la necesidad de calidad de esta cobertura. Así la tecnología 5G se plantea como una alternativa para superar estas brechas.

Tema interesante a resolver será como se articula el potenciar el desarrollo de inteligencia artificial a través de acceso a datos y código abierto, con los temas de privacidad y de propiedad intelectual que se quieran asociar a bases de datos y métodos. Parte importante de esta discusión deberá ser abordada desde una perspectiva global, reconociendo el avance en estas temáticas de otros países y su eventual adaptación en nuestro país.

Existen muchas miradas que avizoran escenarios futuros donde la inteligencia provoca cambios mayores en la sociedad y en la forma en que interactuamos. La inteligencia artificial permite abordar tareas donde distintos tipos de restricciones físicas no permiten su realización por seres humanos. Esto abre las posibilidades a nuevas funcionalidades aún desconocidas, pero además abre la posibilidad de que aumenten las brechas e inequidades entre aquellos que tengan o no tengan el poder de esta tecnología. Enfrentar estos desafíos requiere como paso inicial conocerla, eliminar los mitos y verla como una tecnología que ha sido creada por la inteligencia humana y que debe estar al servicio de la sociedad.

Que la política nacional de inteligencia artificial tenga impacto dependerá no sólo de los recursos que el sector público y privado puedan aportar, sino del compromiso y convicción transversal que vaya más allá de un gobierno para impulsar esta tecnología como agente de desarrollo económico, social y cultural del país. Esto se ve más importante de resolver dada la contingencia de la pandemia, la que ha incentivado la transformación digital, pero que también ha tenido un fuerte impacto económico que hace prever la falta de recursos desde el sector público. Acorde a la trayectoria del aporte del Estado a la investigación en Chile, uno puede pronosticar que las prioridades no irán por apostar al desarrollo científico del área con financiamiento público, sino más bien a la formación de capital humano y al fomento de la innovación con la participación del sector privado. ▪

Sistemas de toma de decisiones automatizadas:

¿De qué hablamos cuando hablamos de transparencia y del derecho a una explicación?

CATHERINE MUÑOZ

Abogada, Magíster en Derecho Internacional, Inversiones y Comercio por la Universidad de Chile y Master of Laws in International Law (LL.M.) por la Universidad de Heidelberg, especializada en propiedad intelectual y regulación de tecnologías, en particular, regulación de inteligencia artificial.

cmunozgut@gmail.com

JEANNA NEEFE MATTHEWS

Profesora de informática en Clarkson University (EE.UU.), copresidenta fundadora del Subcomité de Políticas de Tecnología de la ACM sobre Inteligencia Artificial y Responsabilidad Algorítmica, vicepresidenta del Instituto de Ingenieros Eléctricos y Electrónicos (IEEE) - Comité de Política de IA de EE. UU. y miembro del Comité de Políticas de Tecnología de la ACM (ACM TPC).

jnm@clarkson.edu

JORGE PÉREZ

Profesor Asociado del Departamento de Ciencias de la Computación de la Universidad de Chile e Investigador Asociado del Instituto Milenio Fundamentos de los Datos. Doctor en Ciencias de la Ingeniería por la Pontificia Universidad Católica de Chile. Sus intereses incluyen: datos Web, teoría de redes neuronales profundas, y el análisis de texto en medicina y política. En Twitter lo encuentras como @perez.

Introducción

A mediados de enero de 2021, en un hecho histórico, el Gobierno de los Países Bajos dimitió en bloque luego de una investigación realizada por el parlamento de dicho país que concluyó que el Jefe de Estado y sus principales ministros habían incurrido en faltas graves, evidenciando un menoscabo institucional y una discriminación sistemática contra un grupo vulnerable de la población holandesa. Esta imputación, tiene como fundamento la masiva y errónea acusación de fraude en la obtención de subsidios sociales en contra de 26.000 familias inocentes, de origen marroquí y tunecino en su gran mayoría [1].

La referida investigación constató que un sistema automatizado de toma de decisiones definía aquellos casos sospechosos de fraude en base a variables arbitrarias y abiertamente discriminatorias, como el simple hecho de tener una doble nacionalidad, evento que, por sí solo, situaba a las personas en una categoría de alto riesgo delictual. Lo anterior, unido a una mala gestión administrativa, injustamente obligó a estas familias a devolver dinero de subsidios recibidos. Muchas personas fueron llevadas a la quiebra, otras familias se desintegraron y la gran mayoría padeció estrés psicológico [2].

Lamentablemente este caso no es una excepción. Por el contrario, corresponde a una progresiva e instaurada regla general sobre el uso de sistemas automatizados de toma de decisiones que pueden afectar de manera radical la vida de las personas. Algunos ejemplos incluyen a sistemas predictivos de obtención de beneficios sociales cuya optimización se basó en reducir costos y reducir la mayor cantidad de otorgamiento de beneficios [3], sistemas calificadores de riesgos que utilizaron bases de datos, muchos de ellos con contenido de carácter sensibles, incompletos o falsos, proveídas por empresas *Data Brokers*, sin ningún estándar ético o legal [4], sistemas predictivos de justicia penal que castiga en mayor medida a grupos marginados de la población [5], sistemas de reconocimiento facial sesgados usados con fines de vigilancia y riesgosos resultados erróneos [6], y finalmente, la grave vulneración de derechos humanos, y en particular de la autonomía y privacidad de las personas, derivada del sistema automatizado de calificación de crédito social que impera en China [7].

La implementación de este tipo de sistemas en países en vías de desarrollo, como Chile, evidencian, asimismo, un creciente interés. Chile ha formulado dentro de sus políticas públicas y como

meta a corto plazo, la modernización de sus funciones y prestaciones de servicios [8], incorporando las referidas toma de decisiones automatizadas potenciadas con Inteligencia Artificial (IA). Lo anterior, bajo la consigna de eficiencia pública, administración efectiva y con la promesa de minimizar pérdidas de gastos fiscales, contribuyendo a políticas de austeridad [9].

Desde el punto de vista técnico, los sistemas de tomas de decisiones automatizadas pueden ser, o bien sistemas que apoyan determinadas decisiones teniendo la última palabra un ser humano, o sistemas que toman decisiones sin la intervención de personas [10]. Esta diferencia que pareciese ser trascendental, no es tal y en ambos casos existen similares niveles de riesgos en relación con la afectación de grupos protegidos. Por ejemplo, en el primer caso, también llamado "semiautomatizado", existe una tendencia comprobada; las personas confían más en el juicio de un algoritmo que en el propio cuando estos juicios están en contradicción [3].

Llama la atención que el entusiasmo por este tipo de tecnología no ha mermado a pesar de la abundante evidencia que alerta sobre el riesgo de aplicarlos a problemáticas sociales [11]. El denominador común en su aplicación es la naturaleza punitiva, lo que convierte a estos sistemas en una amenaza potencial de amplificación y perpetuación de injusticias sociales sobre grupos históricamente oprimidos y marginalizados, tales como pueblos originarios, afroamericanos, latinos, asiáticos, comunidades LGBTIQ+, musulmanes, personas de escasos recursos, entre otros [12].

Muchos de estos casos son evidentes e incuestionables discriminaciones, las que legalmente pueden ser acreditadas en un juicio. La información para documentar este tipo de casos toma como referencia los resultados de salida del sistema, junto con pruebas estadísticas y antecedentes relacionados con las personas involucradas en su diseño e implementación, sin necesitar información detallada del funcionamiento interno de los sistemas involucrados. Lo que se busca probar, en estos casos evidentes, es generalmente una discriminación indirecta, la cual ocurre cuando una norma, en este caso un sistema, aparentemente neutro, es aplicado a una población, perjudicando desproporcionadamente a grupos vulnerables de ésta [13]. En consecuencia, la recopilación de este tipo de información, en general, es suficiente para probar dicho "perjuicio dzesproporcionado". Éste es un tipo de "transparencia", pero no cualquiera, sino aquella estratégicamente obtenida para construir un caso judicial donde existe una evidente vulneración de derechos sobre las personas [14].

Ahora bien, ¿qué ocurre en aquellos casos donde la falta, error o injusticia son sutiles y no evidentes? Pensemos en un sistema de contratación de personal que ha rechazado una solicitud de empleo de una persona que cumplía todos los requisitos o un sistema de toma de decisiones que rechaza el ingreso de un joven a una universidad cumpliendo, asimismo, todos los requisitos para ello. Estas personas pueden albergar razonables dudas sobre si han sido injustamente excluidas o discriminadas, pero a diferencia de los casos anteriores, no es algo manifiesto. Incluso más, es posible que estos sistemas ya cuenten con auditorías que demuestren que su funcionamiento está supuestamente libre de sesgos de acuerdo con parámetros matemáticos de equidad [15]. Lamentablemente es común que estos parámetros obedezcan a una visión exclusivamente tecnocrática del problema y tengan poco sustento comparado con parámetros sociales de equidad [16, 17].

Los ejemplos más sutiles de sesgo son muy comunes, lo que va en contra de la creencia de muchas personas de que las decisiones tomadas por computadoras o sistemas automatizados son fundamentalmente lógicas e insesgadas. Y esto no es así. Las decisiones automatizadas se toman de dos formas principales: 1) según las instrucciones escritas por programadores humanos, o 2) según las reglas aprendidas automáticamente a partir de datos del pasado. Algunas personas pueden pensar que el problema principal proviene de las instrucciones escritas directamente por programadores humanos, pero de hecho, el aprendizaje automático sobre datos pasados suele crear problemas aún mayores. Aprender automáticamente desde datos del pasado es equivalente a considerar al pasado como el oráculo del futuro que queremos. En cierto sentido, aprendemos del pasado porque es todo lo que tenemos para aprender. Pero el pasado está lleno de prejuicios de muchos tipos. Si, por ejemplo, miramos quién ha sido un buen gerente en el pasado para definir quién será un buen gerente en el futuro, o quién ha sido un buen enfermero en el pasado para definir quién será un buen enfermero en el futuro, es posible que descartemos personas calificadas que no coinciden con el perfil más típico del pasado. Si codificamos estos datos del pasado en sistemas informáticos sin exigir una explicación de sus decisiones, entonces permitiremos que el pasado defina el futuro sin cuestionarlo. Estaríamos tomando la IA, que consideramos una fuerza progresista y futurista, para usarla como un oráculo y ejecutor conservador de prejuicios pasados.

Los conceptos clásicos de transparencia y participación social en la toma de decisiones, pilares fundamentales para prevenir y combatir la arbitrariedad y la discriminación, parecen quedarse cortos en el contexto actual. En particular, la transparencia puede tener diversas conceptualizaciones y se hace imprescindible distinguir en palabras sencillas transparencia, explicabilidad e interpretabilidad que son términos relacionados mas no sinónimos. ¿Qué exigimos entonces cuando exigimos transparencia y explicabilidad en las decisiones de un sistema automático?

Aprender automáticamente desde datos del pasado es equivalente a considerar al pasado como el oráculo del futuro que queremos.

No pretendemos responder cabalmente a la pregunta sino más bien aportar a la discusión desde una visión legal y computacional. Éste es el punto de partida de este artículo y nuestra motivación de escribirlo.

El concepto clásico de transparencia

Durante la última década se ha discutido sobre el nivel de transparencia que debe existir en el desarrollo y uso de sistemas de IA, en particular, en aquellos que toman decisiones automatizadas y que potencialmente pueden tener un impacto negativo sobre las personas. La transparencia ha sido instaurada como uno de los principios esenciales en esta materia y guarda relación con la capacidad de proporcionar información que permita comprender cómo se desarrolla y despliega un sistema de IA [18, 19]. Al respecto, la Iniciativa Global de IEEE sobre Ética de Sistemas Autónomos e Inteligentes ha establecido cuatro condiciones para guiar la confianza informada de los sistemas autónomos e inteligentes: 1) efectividad, 2) competencia, 3) rendición de cuentas y siendo la 4) precisamente la transparencia [20].

La necesidad de transparencia es contrastada con el hecho de que los sistemas de IA, particularmente los modelos de *deep learning* que tienen una estructura compleja, no permiten transparentar completamente su funcionamiento, siendo en muchos casos imposible explicar la construcción y decisiones de éstos, incluso para sus propios desarrolladores, la famosa caja negra. Más aún, una explicación satisfactoria [21] dependerá de la audiencia; algo que pueda ser considerado como una explicación o evidencia clara para un grupo (p.ej., código fuente de un sistema para un desarrollador de software), puede resultar opaco para otro grupo o simples detalles técnicos para un tercer grupo. A pesar de esto, diversos grupos de investigación están actualmente trabajando en proponer mecanismos para una transparencia efectiva y con sentido.

La transparencia no es sinónimo de igualdad

Comúnmente, el análisis de transparencia es *ex-ante* (antes de que el sistema se implemente), y no *ex-post* (después de que el sistema ya esté implementado y tenga un impacto en la vida de las personas). En ese sentido, se entiende que la transparencia y exigencia de información pertinente, es un requisito para la construcción de la confianza entre los ciudadanos y entidades públicas o privadas y los sistemas que éstos proveen de forma previa a su uso, de manera que las personas puedan contar con antecedentes necesarios para tomar la decisión de aceptar con cierta confianza el uso de un modelo algorítmico que puede impactarlo directamente. Pero esto es cierto sólo respecto de una parte de la población, generalmente de clases acomodadas, ya que respecto de personas vulnerables o de escasos recursos, el uso de sistemas tecnológicos en temáticas que les impactan no les es consultado y menos explicado. Hasta cierto punto, exigir y obtener transparencia es un "privilegio", un elemento más que suma e incrementa la desigualdad estructural de nuestra sociedad. En síntesis, a las personas pobres simplemente les imponen sistemas cuyas decisiones pueden afectar sus vidas a largo plazo independientemente de la transparencia.

En efecto, desde orígenes coloniales las personas de escasos recursos no han tenido control sobre su privacidad ni decisiones, en comparación con personas de clases de mayores ingresos. A lo anterior, se agrega el hecho que, debido a segregaciones y desigualdades, existe una brecha de conocimiento en las personas sobre cómo funcionan las herramientas tecnológicas y la forma en que pueden proteger sus derechos. Adicionalmente en muchos casos, la mayoría de las personas no son conscientes que están siendo parte de sistemas tecnológicos ni de los riesgos asociados [22]. Éste es un aspecto crítico que debe ser democratizado mediante mecanismos de inclusión y en consideración a la dignidad de todos los ciudadanos. Como hemos mencionado, una transparencia suficiente para una persona puede no serlo para otra, por lo que deben existir estándares de acceso a la información que consideren el entendimiento de todos los ciudadanos.

La obtención de información se complejiza, tomando en consideración que existen diferentes definiciones contrapuestas sobre conceptos relevantes como igualdad, discriminación y *fairness* [23]. Por ejemplo, dar prioridad a los derechos de los individuos, priorizar el bienestar de la sociedad en su conjunto, proteger a los grupos marginados, incluso proteger a todas las especies del planeta. *Fairness* es un concepto esencial en países de Europa o en Estados Unidos, que se opone al concepto legal de discriminación, y que posee distintas interpretaciones, dependiendo si se usa en el área computacional, social o legal [24]. Este concepto no posee un equivalente exacto en Chile ni en Latinoamérica, siendo interpretado indistintamente como imparcialidad, equidad o justicia [25] razón por la cual, en este artículo no le daremos una traducción e interpretación determinada.

Dado que las definiciones de *fairness* y ética pueden variar, es especialmente importante que todos los actores que

tienen interés en un sistema, y no sólo los desarrolladores o usuarios contratantes, reciban información que les permita discutir sus prioridades en procesos decisorios. En ese sentido, la transparencia es necesaria para que todas las partes interesadas puedan debatir en un proceso decisorio en torno a la definición de *fairness* que les parezca adecuada y no ceder esta decisión a los creadores, diseñadores y programadores de estos sistemas. En Grasso et al [21] se ha argumentado que el proceso de automatización a menudo desplaza las grandes decisiones de expertos en un dominio determinado hacia programadores sin experiencia en esta área y se discute cómo integrar los marcos de responsabilidad algorítmica con herramientas como "fichas técnicas para *datasets*" [26] y "Tarjetas modelo para informes de modelos" [27] con los códigos de ética específicos de esta materia [21].

La transparencia no es sólo técnica, sino también social

No se debe perder de vista que estamos en presencia de sistemas sociotécnicos. En ese sentido, no pueden ser entendidos sólo desde la técnica, ya que junto a ésta, toman relevancia motivaciones e intereses de las personas que poseen una relación directa en la creación e implementación de un determinado sistema. La suma de factores técnicos y sociales, inciden directamente en los impactos del despliegue de este tipo de tecnología. Como dice Shoshana Zhuboff en su libro The Age of Surveillance Capitalism, debemos preguntarnos: ¿quién sabe? ¿quién decide quién sabe? y ¿quién decide quién decide? [28].

En particular, respecto de sistemas de IA utilizados en políticas públicas, la transparencia desde el punto de vista social se traduce en parte en contar además de información técnica, con información política y social sobre los diseñadores y tomadores de decisiones, sobre la elección de determinados datos, características, modelos, qué tipo de patrones busca, por qué a unas personas sí y otras no, o por qué se dirige a determinado grupo o ámbito geográfico, etc. En definitiva, información sobre las decisiones políticas detrás de las decisiones técnicas.

Para el cumplimiento del estándar anterior, esta transparencia lleva implícita la condición que organismos públicos no adquieran sistemas de IA que estén protegidos por secretos comerciales o acuerdos de confidencialidad. En el mismo sentido, es necesario que exista una transparencia activa del Estado, con mecanismos como registros y plataformas públicas, además de procesos de licitación abiertos. La colaboración público-privada debe ser totalmente transparente, haciendo público conflictos de intereses, contratos con proveedores y cualquier información relevante, cumpliendo con las más altas exigencias de probidad y rendición de cuentas.

Asimismo, en el caso de software de uso público, los gobiernos tienen la oportunidad de establecer requisitos técnicos adicionales tanto para su propio desarrollo como para la compra de software desarrollados por terceros. Así por ejemplo, en la fase de diseño o adquisición se podrían establecer requerimientos de factores pro transparencia, como disponer de software de código abierto, acceso a artefactos de ingeniería de software, incluidos documentos de requisitos y diseño, seguimiento de errores y bitácoras de cambios en el código, planes de prueba y resultados [21].

Explicabilidad e interpretabilidad

Hasta ahora nos hemos concentrado principalmente en el concepto de transparencia de los sistemas automáticos desde una perspectiva general y sobre la necesidad de contar con distintas visiones al momento de su construcción y despliegue.

Hasta cierto punto, exigir y obtener transparencia es un "privilegio", un elemento más que suma e incrementa la desigualdad estructural de nuestra sociedad.

En ese sentido, si bien la transparencia es algo deseable, en la práctica necesitamos también ser capaces de auditar el funcionamiento de los sistemas de manera dinámica, mientras están tomando las decisiones. Es aquí donde surgen dos conceptos que hemos mencionado tangencialmente pero que son de vital importancia: la *interpretabilidad* y la *explicabilidad* de un sistema de toma de decisiones automatizada.

Para una conceptualización útil de explicabilidad, podemos centrarnos en la decisión de un sistema en un caso específico, por ejemplo "una solicitud de crédito que fue rechazada". Lo que buscamos entonces, es que un humano sea capaz de entender la razón de esa decisión particular ("¿por qué fue rechazada la solicitud?"). Usualmente a esto se le llama explicación *post-hoc* y local. *Post-hoc* se refiere a que la explicación se hace considerando los veredictos del sistema después de que el sistema ya está desplegado y en funcionamiento, mientras que local se refiere a explicar una decisión particular (en oposición a explicar el sistema como un todo). Que una decisión sea explicable en un sistema, no significa que el funcionamiento en general (para todas las posibles decisiones) sea explicable también. A esta explicación global le llamamos interpretabilidad; un sistema sería interpretable entonces, si un humano es capaz de entender la manera en que el sistema toma todas sus decisiones.

De la misma manera, se debe tener presente que cualquier explicación es una simplificación del sistema completo. Larraraju et al. [29] establecen claras métricas para determinar la calidad de las explicaciones, que incluyen la fidelidad, es decir, el grado en que la explicación coincide con el sistema completo, la falta de ambigüedad o el grado en que la explicación aísla un único resultado para cada caso, y la interpretabilidad, es decir, el grado en que las personas pueden entender la explicación. La fidelidad puede medirse minimizando la cantidad de desacuerdo entre la explicación y el sistema completo. La falta de ambigüedad puede medirse minimizando la cantidad de solapamiento entre las reglas de la explicación y maximizando el número de casos cubiertos por la explicación. La interpretabilidad puede medirse minimizando el número de reglas, el número de predicados utilizados en esas reglas y la amplitud del número de casos considerados por cada nivel en el árbol de decisiones (por ejemplo, si X_1 entonces Y_1, si X_2 entonces Y_2, si X_3 entonces Y_3, sería de amplitud 3). Otras propiedades deseables de las explicaciones pueden ser que no utilicen características inaceptables (por ejemplo, utilizar la raza o el género en las decisiones de contratación) o que proporcionen una orientación predictiva (por ejemplo, si tuviera más experiencia en la categoría *X*, tendría más probabilidades de ser contratado para este trabajo en el futuro). En definitiva, en la comunidad científica se sigue trabajando en las características de las buenas explicaciones y existe una tensión natural entre diferentes características como la interpretabilidad y la fidelidad, aún no resuelta.

Un intento de formalización y la esperanza de auditabilidad

La anterior discusión se basa en que "un humano sea capaz de entender" algo, lo que es sumamente difícil de formalizar y definir de una única forma. Una manera de concretizar el problema es llevarlo a un tipo particular de explicación. Una muy usada es la del tipo contrafactual; en vez de preguntarnos el porqué de la decisión, nos preguntamos cómo cambiaría la decisión en presencia de antecedentes distintos ("¿hubiese sido rechazada la solicitud si el postulante hubiera sido una persona casada?"). Este tipo de preguntas se han usado recientemente para comparar la interpretabilidad de distintos sistemas de manera formal independiente de las características del sistema en cuestión. Más precisamente, supongamos que un sistema *M* toma cierto veredicto cuando es presentado con un conjunto *A* de antecedentes, y consideremos la siguiente pregunta: ¿cuál es el mínimo grupo de antecedentes que es necesario cambiar en *A* para cambiar también el veredicto de *M*? Podríamos definir entonces que un sistema automático es interpretable, si para cada posible conjunto de antecedentes, la anterior pregunta se puede responder en un tiempo prudente ("tiempo polinomial" en jerga computacional). Esta definición aseguraría que, por ejemplo, cada persona a la que se le haya rechazado una solicitud de crédito, podría obtener en un tiempo prudente una explicación del tipo "si cambia este grupo de antecedentes, el crédito sería aprobado".

Sin perjuicio de lo anterior, debemos notar que esta definición de interpretabilidad es sumamente acotada y posiblemente sea útil sólo en ciertos contextos. Si bien esta perspectiva es acotada, es formal, y una de las consecuencias de definir formalmente un problema de interpretabilidad, es que podemos poner a prueba de manera precisa y comparativa

Existe el riesgo de que los usuarios entendamos la explicación [acerca de la respuesta otorgada por un sistema automático] como producto de causalidades.

a distintas clases de sistemas automatizados. En efecto, con esta definición se puede demostrar formalmente la creencia popular de que sistemas basados en árboles de decisión son más interpretables que sistemas basados en redes neuronales profundas [30, 31]. Otro punto positivo de contar con una definición del tipo anterior, es que un sistema interpretable se podría auditar respecto de la existencia de sesgos en sus veredictos. Por ejemplo, si hubiese un conjunto de antecedentes protegidos (como género o raza), podríamos exigir de manera efectiva que el solo cambio de esos antecedentes protegidos no cambien el veredicto del sistema [32].

Si bien hemos mostrado posibilidades de resolver problemas de interpretabilidad de una manera un poco más precisa, la aplicación de la definición anterior (o cualquier otra que se proponga desde la técnica), no debiera obviar aspectos sociales. Por ejemplo, no debieran ser los mismos sistemas los que definan cuáles son los antecedentes protegidos. También se debe tomar en cuenta que las explicaciones serán consumidas por personas y por lo tanto se debiera evitar la jerga técnica y presentar explicaciones precisas pero simples de entender, que incluyan modelos cuantitativos, cualitativos y antropológicos, entre otros [33].

Explicaciones *post-hoc*, locales, basadas en contrafactuales y que puedan generarse en tiempo razonable (polinomial), son esencialmente conceptos técnicos y las formalizaciones han venido principalmente desde el mundo científico. En consecuencia, no debemos perder de vista que cualquier definición técnica puede tener implicancias en la forma en que las personas entenderán el proceso real para el que se usa el sistema. Por ejemplo, una explicación contrafactual ("qué habría pasado si cambiaba el antecedente *x*") no es necesariamente causal ("el antecedente *x* es el más importante en la decisión del sistema") sin embargo existe el riesgo de que los usuarios entendamos la explicación como producto de causalidades [34]. Se hace necesario entonces que la sociedad, y más precisamente la legislación, defina, al menos conceptualmente, qué tipo de explicaciones, interpretaciones y estándares deben ser exigidos a los sistemas automatizados. Visualizamos acá un círculo virtuoso: las definiciones sociales podrán guiar el desarrollo técnico, incentivando la cooperación y búsquedas de soluciones interdisciplinarias, enfocando así recursos y esfuerzos de investigación.

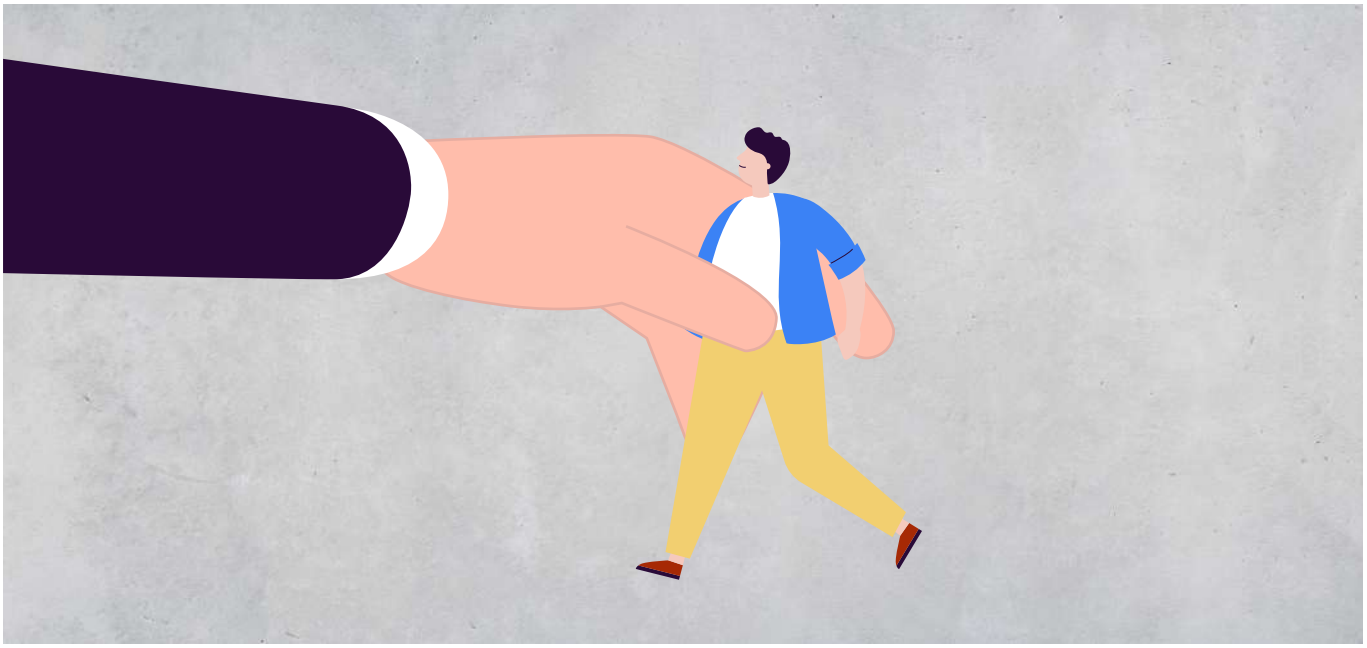
Transparencia algorítmica y el proceso constituyente

Mientras en todo el mundo los sistemas basados en IA están cambiando la forma en que se deciden aspectos importantes de la vida de las personas, Chile se encuentra en un proceso histórico de diseño de una nueva Constitución. En este contexto, Chile tiene la oportunidad de delinear el rol que los sistemas de IA tendrán en la toma de decisiones acerca de la asignación de fondos públicos, puestos de trabajo, vivienda, créditos, acceso a la salud, justicia, prevención del delito y muchos otros.

La transparencia, como un concepto general, más que un principio propiamente tal, es un medio que hace posible lograr el ejercicio de derechos fundamentales. Esto toma una relevancia adicional en relación con el uso de sistemas de IA. La Constitución, además de mantener el equilibrio de los poderes del Estado, consagra derechos fundamentales. De estos derechos, los que más riesgo de vulnerabilidad corren a la luz del uso de sistemas de tomas de decisiones automatizadas poco transparentes o no explicables, corresponden principalmente a los derechos de igualdad, privacidad y protección de datos, debido proceso y acceso a un juicio justo, seguridad, autonomía, así como, acceso a información y libertad de expresión.

Respecto del derecho de igualdad consideramos que es una oportunidad histórica consagrar expresamente a la igualdad no como "no discriminación" sino como un principio de antisubordinación. El propósito del principio de igualdad desde esta perspectiva (que muchos autores llaman igualdad real) tiene por finalidad eliminar las estructuras sociales históricamente discriminatorias y excluyentes [35]. Lo anterior tiene una importante consecuencia sobre la regulación de sistemas de toma de decisiones automatizadas, ya que se traduce en que cualquier resultado de éstos, que reproduzca y perpetúe condiciones estructurantes de injusticia social, no serán tolerados por la legislación y serán sancionados, sin considerar otros elementos como la intención de provocar daños. Este punto es importante cuando no podemos contar con toda la transparencia requerida frente a potenciales efectos negativos en el uso de sistemas de toma de decisiones automatizadas.

Por su parte, sobre la protección de la privacidad y la protección de datos personales, la transparencia, y la interpretabilidad, cumplen un rol fundamental. Notable es el caso de los artículos 13° y 15° del Reglamento General de Protección de Datos (GDPR, por sus siglas en inglés) en Europa, que proveen el derecho a una "explicación significativa de



la lógica involucrada" en las decisiones automáticas. Selbst y Powles [36] consideran que esto traza un fundamento claro hacia el "derecho a la explicación", que son complementadas con los artículos 22° y 35° del mismo cuerpo legal. Chile tiene una oportunidad histórica de consagrar de manera no ambigua en su nueva Constitución el "derecho a la explicación" respecto de sistemas de IA, en particular, de toma de decisiones automatizadas.

Considerando lo descrito en puntos anteriores, específicamente sobre los límites y riesgos de explicaciones descontextualizadas o no entendidas, creemos que tomando todas las prevenciones del caso, es fundamental el establecimiento de un "Derecho a la transparencia y suministro de información sobre sistemas de toma de decisiones automatizadas", consagrados en la nueva Constitución dentro de un "Derecho a la transparencia e información" de carácter más general, el cual para garantizarlo, debe ser complementado con la promulgación de normas de rango legal en donde se detallen los mecanismos y estándares para su cumplimiento. Al respecto, la reciente publicación de la Propuesta de Reglamento del Parlamento Europeo y del Consejo Europeo que establece normas armonizadas sobre la inteligencia artificial (Ley de Inteligencia Artificial, publicada con fecha 21 de abril de 2021 [EU Council 2021]), es un excelente ejemplo del contenido mínimo que debieran tener estas futuras normas legales, además de las ya referidas al GDPR, para el debido ejercicio de este nuevo derecho constitucional.

La Propuesta de Reglamento del Parlamento Europeo sobre la inteligencia artificial establece estándares de transparencia, registro y explicabilidad, respecto de sistemas considerados por este cuerpo legal como de alto riesgo, y que pueden ser resumidos en los siguientes puntos:

a. Deben contener instrucciones de uso con información concisa, pertinente, accesible y comprensible, sobre datos de proveedor, características, capacidades y limitaciones de funcionamiento, finalidad prevista, rendimiento, especificaciones de los datos de entrada, las medidas de supervisión humana, incluidas las medidas técnicas establecidas para facilitar la interpretación de los resultados de los sistemas de IA por parte de los usuarios; entre otros.

b. Deben contener documentación técnica sobre finalidad prevista, desarrolladores, la interacción del sistema con hardware o software que no forma parte del mismo, los métodos y pasos realizados para el desarrollo del sistema, incluido, el uso de sistemas preentrenados o de herramientas proporcionadas por terceros, lógica general del sistema y de los algoritmos, las opciones clave de diseño, las personas o grupos de personas con los que se pretende utilizar el sistema, opciones de clasificación, entre otras.

c. Información detallada sobre el seguimiento, el funcionamiento y el control de sistemas de IA, en particular, respecto a sus capacidades y limitaciones, incluidos los grados de precisión para grupos de personas específicos en los que se prevé utilizar y el nivel general de precisión esperado en relación con su finalidad prevista. A este último punto se debe complementar el requisito que el nivel de precisión debe estar avalado por metodologías con bases científicas robustas e independientes.

A lo anterior, se debiese agregar la obligación de efectuar una evaluación de impacto en relación con la afectación de derechos humanos. Las evaluaciones dejan documentado el proceso de acuerdo con la letra (b) precedente y permiten prever riesgos antes de su implementación y posibles mejoras o derechamente decidir sobre su no uso.

Conclusiones

La transparencia y el acceso a la información es una idea que ha ocupado un lugar destacado en la agenda política de las sociedades democráticas occidentales durante muchos años. Ha sido cultivada, propagada y, a veces, mal utilizada por los medios de comunicación en forma interesada.

En este artículo intentamos contribuir a la discusión, considerando la importancia de distinguir las distintas funciones de la transparencia y de contar con explicaciones e interpretaciones sobre las decisiones que toman los sistemas automáticos de manera que todas las partes interesadas y posiblemente afectadas puedan entender y responder a ellas.

En particular, consideramos que se debe promover un acceso equitativo sobre transparencia social y aspectos técnicos, teniendo presente que estamos frente a sistemas sociotécnicos, así como promover el acceso a información interpretable que pueda ser usada por profesionales especializados. Para ello nos encontramos en una oportunidad histórica de plasmarlo en nuestra nueva Constitución como un derecho consagrado para todos los chilenos.

Lo anterior en ningún caso se debe interpretar como que estas propuestas conllevan una carga sobre las personas respecto de la decisión de determinar si un sistema de IA es confiable o no. Sería una carga injusta para lo cual no estamos capacitados, por lo que siempre será una obligación del Estado asegurar que estos sistemas sean confiables y cumplan con todos los estándares necesarios para la protección de los ciudadanos y en particular de aquellos más vulnerables.

Finalmente, tanto o más importante que decidir qué rol esperamos que cumpla la IA, es el determinar qué rol esperamos que no cumpla y para ello el análisis en el uso de sistema de toma de decisiones automatizadas no puede ser abordado netamente desde una perspectiva económica de costos versus beneficios, sino que se debe considerar si corresponde desplegar este tipo de sistemas en consideración a los derechos y dignidad de las personas. Para asegurarnos de que esto se cumpla, requerimos, nuevamente, transparencia e información.

Como profesionales del área legal y de las ciencias de la computación, sabemos que los sistemas computacionales complejos cometen errores, y a veces muchos errores. Por eso estamos en contra de un mundo regido por el principio de que "el computador sabe más que nadie" o la creencia de que, a diferencia de los humanos, los sistemas automáticos "pueden tomar decisiones sin sesgos". Soluciones simplistas, o que sólo vengan del mundo técnico podrían, más que ayudar, crear más daño. Éste es uno de esos problemas en donde basados en ciencia y evidencia, pero sobre todo basados en el bien común, debemos buscar una solución como sociedad. ▪

REFERENCIAS

[1] G. Geiger, «How a Discriminatory Algorithm Wrongly Accused Thousands of Families of Fraud», ene. 01, 2021. https://www.vice.com/en/article/jgq35d/how-a-discriminatory-algorithm-wrongly-accused-thousands-of-families-of-fraud (accedido abr. 28, 2021).

[2] T. K. der Staten-Generaal, «Parlementaire ondervraging kinderopvangtoeslag; Brief Presidium; Brief van het Presidium over een voorstel voor een parlementaire ondervraging kinderopvangtoeslag», jul. 01, 2020. https://zoek.officielebekendmakingen.nl/kst-35510-1 (accedido abr. 28, 2021).

[3] V. Eubanks, *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press, 2018.

[4] H. Fry, *Hello world: Being human in the age of algorithms*. WW Norton & Company, 2018.

[5] J. N. Matthews *et al.*, «When Trusted Black Boxes Don't Agree: Incentivizing Iterative Improvement and Accountability in Critical Software Systems», 2020, pp. 102-108.

[6] K. Hill, «What Happens When Our Faces Are Tracked Everywhere We Go?», *The New York Times*, mar. 18, 2021.

[7] S. Engelmann, M. Chen, F. Fischer, C.-Y. Kao, y J. Grossklags, «Clear Sanctions, Vague Rewards: How China's Social Credit System Currently Defines "Good" and "Bad" Behavior», ene. 2019, pp. 69-78, doi: 10.1145/3287560.3287585.

[8] https://digital.gob.cl, «Ley de Transformación Digital», *Ley de Transformación Digital*. http://digital.gob.cl/transformacion-digital/ley-de-transformacion-digital/ (accedido abr. 28, 2021).

[9] J. Hughes, «Algorithms and posthuman governance», *J. Posthuman Stud.*, vol. 1, n.º 2, pp. 166-184, 2018.

[10] C. Orwat, «Risks of Discrimination through the Use of Algorithms. A study compiled with a grant from the Federal Anti-Discrimination Agency», 2020.

[11] F. Chiusi *et al.*, «Automating Society Report 2020», *Automating Society Report 2020*. https://automatingsociety.algorithmwatch.org (accedido abr. 28, 2021).

[12] R. Benjamin, «Race after technology: Abolitionist tools for the new jim code», *Soc. Forces*, 2019.

[13] T. Khaitan, *A theory of discrimination law*. OUP Oxford, 2015.

[14] S. Wachter, B. Mittelstadt, y C. Russell, «Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI», *ArXiv Prepr. ArXiv200505906*, 2020.

[15] K. Creel y D. Hellman, «The Algorithmic Leviathan: Arbitrariness, Fairness, and Opportunity in Algorithmic Decision Making Systems», *Va. Public Law Leg. Theory Res. Pap.*, n.o 2021-13, 2021.

[16] A. D. Selbst, D. Boyd, S. A. Friedler, S. Venkatasubramanian, y J. Vertesi, «Fairness and abstraction in sociotechnical systems», 2019, pp. 59-68.

[17] M. Srivastava, H. Heidari, y A. Krause, «Mathematical notions vs. human perception of fairness: A descriptive approach to fairness for machine learning», 2019, pp. 2459-2468.

[18] S. Garfinkel, J. Matthews, S. S. Shapiro, y J. M. Smith, «Toward algorithmic transparency and accountability», 2017.

[19] A. Now, «The Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems», *https://www.accessnow.org/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems/*, 2018.

[20] K. Shahriari y M. Shahriari, «IEEE standard review—Ethically aligned design: A vision for prioritizing human wellbeing with artificial intelligence and autonomous systems», 2017, pp. 197-201.

[21] I. Grasso, D. Russell, A. Matthews, J. Matthews, y N. R. Record, «Applying Algorithmic Accountability Frameworks with Domain-specific Codes of Ethics: A Case Study in Ecosystem Forecasting for Shellfish Toxicity in the Gulf of Maine», 2020, pp. 83-91.

[22] M. Madden, M. Gilman, K. Levy, y A. Marwick, «Privacy, poverty, and big data: A matrix of vulnerabilities for poor Americans», *Wash UL Rev*, vol. 95, p. 53, 2017.

[23] A. Narayanan, «Translation tutorial: 21 fairness definitions and their politics», 2018, vol. 2, n.o 3, pp. 6-2.

[24] A. Xiang y I. D. Raji, «On the legal compatibility of fairness definitions», *ArXiv Prepr. ArXiv191200761*, 2019.

[25] J. Rawls, «Justice as fairness», *Philos. Rev.*, vol. 67, n.o 2, pp. 164-194, 1958.

[26] T. Gebru *et al.*, «Datasheets for datasets», *ArXiv Prepr. ArXiv180309010*, 2018.

[27] M. Mitchell *et al.*, «Model cards for model reporting», 2019, pp. 220-229.

[28] S. Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power: Barack Obama's Books of 2019*. Profile Books, 2019.

[29] H. Lakkaraju, E. Kamar, R. Caruana, y J. Leskovec, «Faithful and customizable explanations of black box models», 2019, pp. 131-138.

[30] P. Barceló, M. Monet, J. Pérez, y B. Subercaseaux, «Model Interpretability through the lens of Computational Complexity», *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 15487-15498, 2020.

[31] Z. C. Lipton, «The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery.», *Queue*, vol. 16, n.o 3, pp. 31-57, jun. 2018, doi: 10.1145/3236386.3241340.

[32] P. Barceló, J. Pérez, y B. Subercaseaux, «Foundations of Languages for Interpretability and Bias Detection». *Algorithmic Fairness through the Lens of Causality and Interpretability Workshop at NeurIPS 2020*

[33] M. M. Malik, «A Hierarchy of Limitations in Machine Learning», *ArXiv Prepr. ArXiv200205193*, 2020.

[34] R. Moraffah, M. Karami, R. Guo, A. Raglin, y H. Liu, «Causal interpretability for machine learning-problems, methods and evaluation», *ACM SIGKDD Explor. Newsl.*, vol. 22, n.o 1, pp. 18-33, 2020.

[35] R. B. Siegel, «Equality talk: Antisubordination and anticlassification values in constitutional struggles over Brown», *Harv Rev*, vol. 117, p. 1470, 2003.

[36] A. D. Selbst y J. Powles, «Meaningful Information and the Right to Explanation», Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 3039125, nov. 2017. Accedido: abr. 28, 2021. [En línea]. Disponible en: https://papers.ssrn.com/abstract=3039125.

Una dicotomía engañosa y una paradoja ética

RICARDO BAEZA-YATES

Profesor de Investigación del Instituto de Inteligencia Artificial Experiencial de Northeastern University, además de Profesor Titular a tiempo parcial en los Departamentos de Tecnologías de la Información y de las Comunicaciones de la Universitat Pompeu Fabra en Barcelona y Ciencias de la Computación de la Universidad de Chile, donde además es Investigador Senior del Instituto Milenio Fundamentos de los Datos. Entre 2006 y 2016, fue vicepresidente de investigación de Yahoo! Labs, primero desde Barcelona y luego en Sunnyvale, California. Es ACM e IEEE Fellow. En Twitter lo encuentras como @PolarBearby.

Hace poco más de un año, el 20/02/2020 —¿fecha aleatoria?—, Geoff Hinton, uno de los padres del aprendizaje profundo, tuiteó lo siguiente [1]:

> *Suponga que tiene cáncer y tiene que elegir entre un cirujano de IA de caja negra que no puede explicar cómo funciona, pero tiene una tasa de éxito del 90% y un cirujano humano con una tasa del 80%. ¿Quiere que el cirujano de IA sea ilegal?*

Esta provocativa (doble) pregunta incluye una dicotomía engañosa, ya que una persona racional no decidiría basándose sólo en un promedio que ni siquiera sabe cómo se calculó. Engañosa porque hay una tercera posibilidad que es mucho mejor: *quiero un cirujano humano con apoyo de IA*. Pero el verdadero dilema está en la legalidad del cirujano de IA, lo cual es confuso, porque si fuera ilegal, la elección sería retórica.

Los comentarios en este tuit son en su mayoría en contra del cirujano de IA por muchas razones, la mayoría de las cuales se incluyen a continuación. Para empezar, supongamos que el robot médico no es ilegal y sólo tenemos estas dos opciones. ¿Cuál es el mejor? Dando el beneficio de la duda al robot médico, la respuesta depende de la *complejidad del cáncer* y la *normalidad del paciente*, es decir qué tan diferente es usted de la población usada en los datos de entrenamiento. Entonces, si conocemos estos datos y usted es un caso estándar, puede elegir con seguridad al cirujano de IA. En todos los demás casos, es mejor seguir con un médico humano que pueda generalizar y lidiar con problemas inesperados en función de su experiencia. El diagrama de la Figura 1 muestra estas dos dimensiones.

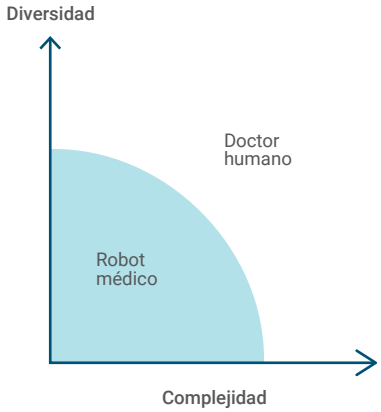


Figura 1. Preferencia de un doctor humano o un robot médico, según la complejidad del cáncer (eje *x*) y la normalidad del paciente (eje *y*) a operar.

Supongamos que: (1) el robot médico usó *buenos datos de entrenamiento* y aprobó todas las regulaciones legales (por ejemplo, regulación para dispositivos médicos); (2) conocemos la *distribución del error con respecto a la complejidad del caso* (aunque la mayoría de los sistemas de IA ni siquiera reportan el error promedio); y (3) conocemos los *sesgos y debilidades del sistema* con respecto a la diversidad de pacientes (por ejemplo, enfermedades actuales, peso, presión arterial, etc.). Sí, estoy *suponiendo muchas cosas*.

Como casi no existen estudios de distribución del error de predicción en función de la complejidad de la instancia del problema, supondremos que el error relativo es proporcional a la complejidad del caso dividida por el tamaño de los datos de entrenamiento usados para ese nivel de complejidad (ver Figura 2). Esto supone que los datos de entrenamiento son homogéneos, lo que difícilmente es cierto en la práctica, ya que normalmente hay menos datos para instancias más complejas. Observe que estoy usando el error de predicción en un sentido amplio, pero en nuestro caso particular sería el error en la predicción de tener una cirugía exitosa o, en otras palabras, salvar al paciente.

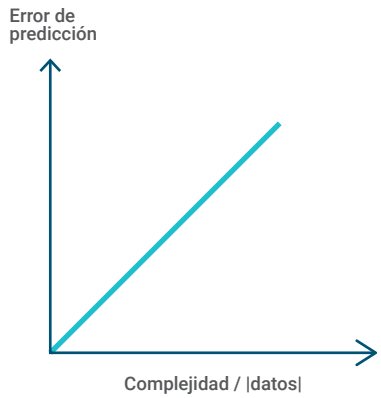


Figura 2. Error en la predicción del éxito de un robot médico, en función de la complejidad de la cirugía y el tamaño de los datos de entrenamiento.

Para el médico humano consideremos que (1) tiene mucha experiencia, lo que le permite transferir sus conocimientos a eventos inesperados en casos complicados,

Para tener mejores robots médicos para todos, debemos correr más riesgos con algunos pacientes, lo que tiene varias facetas éticas [...].

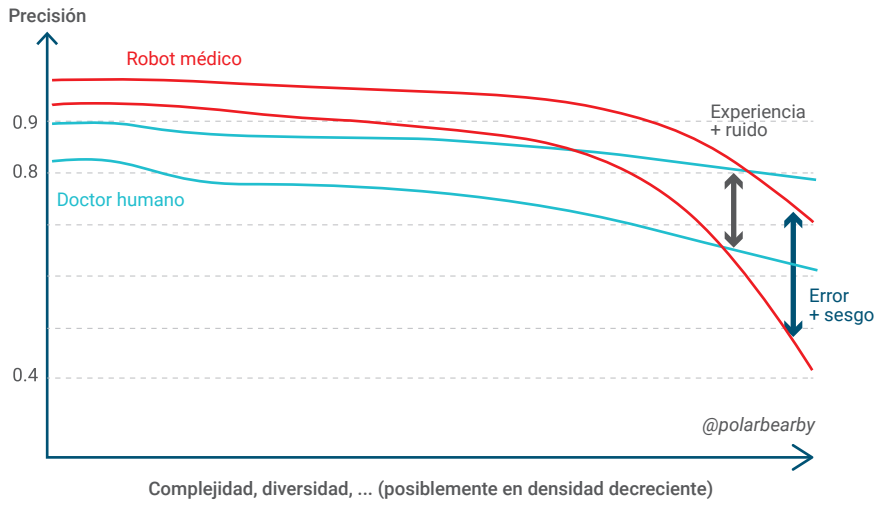


Figura 3. Precisión de éxito de una cirugía (eje *y*) en función de los diversos factores involucrados (eje *x*) para ambos doctores.

así como evitar sesgos de diversidad de la anatomía humana; pero (2) es humano y, por lo tanto, su desempeño varía según su estado de ánimo, las horas de sueño y otros factores. Estos son ejemplos del *ruido que existe en las decisiones humanas* y que tiene un costo tan o más alto que el de los sesgos [2, 3].

Dado lo anterior, la precisión de tener una cirugía exitosa en función de estos factores para ambos médicos viene dada por el gráfico de la Figura 3. Por lo tanto, si sabe dónde usted se encuentra en este gráfico, según la complejidad de su cáncer y lo normal/diferente que usted sea, puede tomar la decisión correcta utilizando el primer diagrama. Es decir, si no coincide con el "caso estándar", mejor elegir un médico humano. Si se encuentra en el lado derecho del gráfico para las dos dimensiones utilizadas en el ejemplo, su riesgo es mayor y es por eso que se obtiene un cuarto de círculo.

Sin embargo, en la práctica no tenemos tanta información. Primero, deberá hablar con el médico humano para saber cuán complejo es su caso, ya que el cirujano de IA no puede explicar nada. Hinton diría que el médico inventará una historia, no una explicación, como en su entrevista en *Wired* [4] que fue refutada por ocho expertos en *Forbes* [5]. Incluso si creemos esto, las personas necesitan historias para sentirse seguras y tomar decisiones importantes. La vida, lamentablemente, incluye también las percepciones además de la realidad, basta con ver lo que está sucediendo con el novel coronavirus.

Ahora, supongamos que el médico humano le dice que su caso es bastante estándar (de lo contrario, ya sabe que probablemente esté más seguro con el médico humano). Ahora necesita conocer las características de los datos de entrenamiento, pero los trabajos que describen la investigación que condujeron al cirujano de IA no son tan detallados y cuando le pregunta al fabricante, le dice que los datos son privados (según las regulaciones legales que el fabricante de cirujanos de IA debe cumplir, no porque no quiera que nadie lo audite, particularmente con respecto al sesgo de diversidad).

En este punto, su única opción es hacer una suposición informada basada en cuán "distinto" es usted. Si es bastante distinto, consulte con el médico humano. Si usted es un caso típico (por ejemplo, un hombre blanco en los Estados Unidos), puede asumir que los datos de entrenamiento representan muchas de esas personas y puede elegir con seguridad al robot médico.

Sin embargo, según la ley de Murphy, usted está en el medio de la región derecha del gráfico, donde la línea roja inferior y la amarilla superior ya se cruzaron. Mala suerte, los trabajos de investigación tampoco tienen la distribución de errores, por lo que realmente no conoce las líneas rojas. Pero, tiene suerte con las líneas amarillas, ya que encuentra un foro con muchas opiniones sobre el médico humano, pudiendo estimar su experiencia y ruido. No ocurre lo mismo con el cirujano de IA porque es bastante nuevo.

Ahora se encuentra en un dilema, ya que sabe que el cirujano de IA tiene una ventaja del 10%. Pero debido al sesgo de diversidad desconocido, la incertidumbre de un error de IA negativo y la fe en un ruido positivo, el médico humano podría ser igualmente bueno. *Piensa esto tres veces y elige al médico humano*. ¿Por qué? Sus instintos se lo dicen. Pero también sabe que el médico puede explicar más tarde lo que sucedió y en el caso de un problema, su familia podría estar psicológicamente mejor, ya que culpar a una máquina no es fácil. Más aún si el robot no puede explicar nada. Además, no ha tenido tiempo de leer todas las cláusulas minúsculas de su seguro médico y tampoco quiere arriesgarse a dejar a su familia con una enorme deuda de salud.

Si los robots médicos no son ilegales, al menos necesitan una regulación estricta con respecto a los datos de entrenamiento, pruebas estándar contra resultados sesgados y algún nivel de explicación.

¿Podemos responder ahora si el cirujano de IA debería ser ilegal? Probablemente debería ser legal pero no es una pregunta sencilla. Una ley de este tipo implica temas éticos, empatía y otros rasgos humanos. Por supuesto, si los robots médicos no son ilegales, al menos necesitan *una regulación estricta con respecto a los datos de entrenamiento, pruebas estándar contra resultados sesgados y algún nivel de explicación*, incluso si necesitan inventar historias. También deben advertirnos cuándo no utilizarlos, ya que tomar la decisión correcta, como hemos visto, no es trivial. Esto es hoy estándar en los medicamentos (por ejemplo, mujeres embarazadas, personas con presión arterial alta, alergias, etc.). Si ese es su caso, ni siquiera tendrá que elegir, el médico humano le dirá directamente que es un caso de riesgo para el robot médico. Sí, quiero este tipo de explicaciones, ¡y mejor si vienen de un doctor empático! (muy probablemente una mujer, un sesgo positivo).

Pareciera estar todo claro, sin embargo, hay una *paradoja ética* escondida en nuestra discusión anterior. Para mejorar el cirujano basado en IA, ellos necesitan aprender y para eso necesitamos tener grandes maestros humanos que puedan generar datos de entrenamiento asombrosos. Para ello, necesitan practicar en los casos más complejos, por lo que ésta es una razón social para preferir cirujanos humanos en los casos más arriesgados. Pero para llegar a este punto, los médicos humanos deben adquirir experiencia en casos estándares, lo que implica que también deben realizar cirugías cuando el cirujano de IA sería una mejor opción. Ésta es la paradoja, para tener mejores robots médicos para todos, debemos correr más riesgos con algunos pacientes, lo que tiene varias facetas éticas que como objetivo final tienen el bienestar común de todos. Lo más justo sería hacerlo al azar, pero no es tan sencillo en un mundo capitalista lleno de sesgos. Éste ya es un tema para filósofos y economistas. ▪

REFERENCIAS

[1] Geoff Hinton, https://twitter.com/geoffreyhinton/status/1230592238490615816, 2/2020.

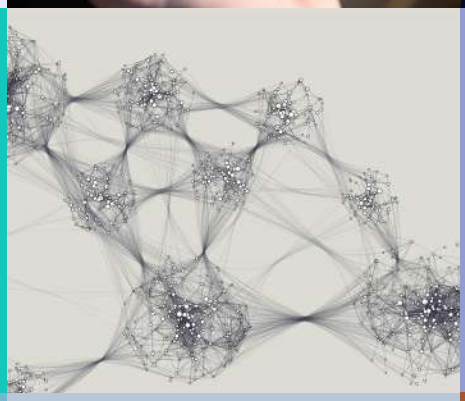
[2] Daniel Kahneman, Andrew M. Rosenfield, Linnea Gandhi, and Tom Blaser Noise: How to Overcome the High, Hidden Cost of Inconsistent Decision Making, Harvard Business Review, https://hbr.org/2016/10/noise, 10/2016.

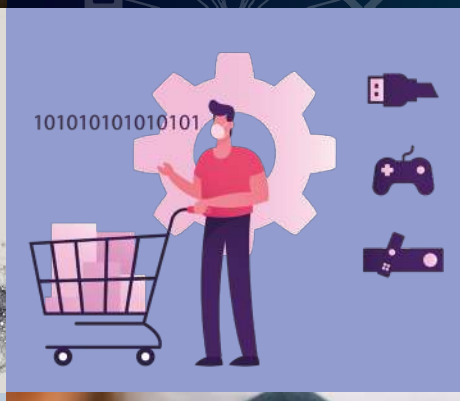
[3] Daniel Kahneman, Olivier Sibony, Cass R. Sunstein. Noise: A Flaw in Human Judgment. Little, Brown Spark, 5/2021.

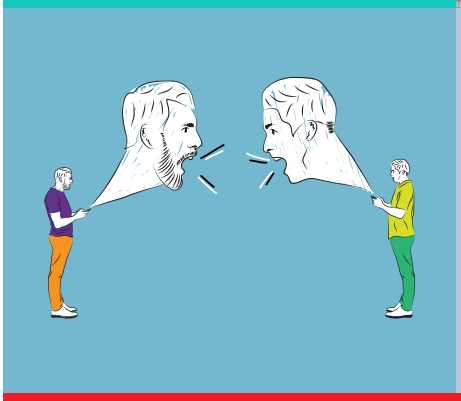
[4] Tom Simonite: Google's AI Guru Wants Computers to Think More Like Brains, Wired, https://www.wired.com/story/googles-ai-guru-computers-think-more-like-brains/, 12/2018.

[5] Hessie Jones: Geoff Hinton Dismissed the Need for Explainable AI: 8 Experts Explain Why He's Wrong, Forbes, https://www.forbes.com/sites/cognitiveworld/2018/12/20/geoff-hinton-dismissed-the-need-for-explainable-ai-8-experts-explain-why-hes-wrong/, 12/2018.

Aplicaciones de la inteligencia artificial







A través de una serie de miniartículos independientes, ilustramos cómo la inteligencia artificial y sus diferentes métodos permiten abordar problemas en una amplia y creciente diversidad de dominios. Por cuestiones de extensión, la enumeración no pretende ser exhaustiva y muchas áreas quedarán pendientes para una futura edición de la Revista.

¿Puede una máquina ver mejor que un humano?

JAVIER CARRASCO Ingeniero Civil en Computación de la Universidad de Chile y egresado del Instituto Milenio Fundamentos de los Datos.

AIDAN HOGAN Profesor Asociado del Departamento de Ciencias de la Computación de la Universidad de Chile e Investigador Asociado del Instituto Milenio Fundamentos de los Datos.

JORGE PÉREZ Profesor Asociado del Departamento de Ciencias de la Computación de la Universidad de Chile e Investigador Asociado del Instituto Milenio Fundamentos de los Datos.

La última década ha sido testigo de avances extraordinarios en el área de la inteligencia artificial, impulsados, en particular, por el concepto de redes neuronales profundas, combinado con la disponibilidad de enormes cantidades de datos para entrenar estas redes. Entre las subáreas de la computación que se han beneficiado con esta tecnología, podemos destacar, por ejemplo, la visión computacional, y la tarea específica de reconocimiento de imágenes. En esta tarea, la máquina recibe una imagen de un objeto y tiene que devolver la clase de ese objeto, diciendo, por ejemplo, que la imagen representa un perro, una flor, una taza, etc.

El conjunto de datos más usado para entrenar y evaluar métodos de reconocimiento de imágenes se llama ImageNet; contiene millones de imágenes etiquetadas según mil clases distintas. Según Russakovsky et al. [1], un experto humano puede lograr una tasa de error (top-5) de 5,1% en un subconjunto de 1.500 imágenes de ImageNet. En la misma tarea, una red neuronal profunda del estado del arte (SeNetResNet50 [2]) puede lograr una tasa de error (top-5) de 2,3%, es decir que tiene mejor rendimiento que un humano experto en esta tarea. ¿Este resultado significa que las máquinas, ahora, pueden "ver" mejor que los humanos? No necesariamente, pues es una pregunta multifacética. En esta tarea, las clases son muy finas, e incluyen ejemplos como un *cucal*, un *Sealyham terrier*, etc., que pueden ser difíciles de recordar y distinguir para un humano. También, la tarea siempre considera imágenes de calidad total. Entonces surge una duda: si las imágenes tuvieran menos calidad que las vistas en los ejemplos de entrenamiento, ¿cómo afectaría el rendimiento de las máquinas y de los humanos? ¿Los humanos necesitan más o menos información para poder clasificar una imagen correctamente en comparación con las máquinas? ¿Qué tipo de información les importa más?

Imágenes mínimas positivas

Para poder entender y comparar la dependencia que las máquinas y los humanos tienen para poder clasificar bien una imagen, definimos el concepto de una *imagen mínima positiva* [3]: dada una imagen etiquetada con su clase, y un clasificador de imágenes, la imagen mínima positiva es la versión de la imagen con la peor calidad tal que el clasificador siga dando la clase correcta. Con respecto a la calidad de la imagen, hablamos más específicamente de

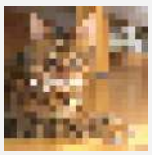
Modelo	Color	Resolución	Zona	Combinación
SqueezeNet				
GoogLeNet				
ResNet50				
SeNetResNet50				
Humano				

Figura 1. Imágenes mínimas positivas para un gato.

la información que contiene, medida usando el tamaño de la imagen comprimida (sin pérdida; usamos compresión de PNG). Se pueden considerar varias formas de reducción de imágenes; en nuestro trabajo, hemos considerado las reducciones de color, de resolución, de zona, y la combinación de las tres. La tabla de la Figura 1 ejemplifica las imágenes mínimas para una imagen de un gato, tal que el modelo (clasificador) indicado puede reconocer que la imagen es de un gato, pero con más reducción, no puede más.

Para calcular las imágenes mínimas en el caso de las máquinas, tomamos una imagen de prueba (no vista antes durante el proceso de entrenamiento), e implementamos una búsqueda sobre los parámetros de reducción, empezando con la imagen completa, y reduciendo la información hasta que se encuentre la imagen mínima. Para calcular las imágenes mínimas en el caso de las máquinas, no se puede usar la misma estrategia, pues el humano recordará la clase de la imagen completa. Así que diseñamos una interfaz que empieza con la imagen "nula" (con una reducción completa), tal que el humano pueda aumentar la información hasta que pueda reconocer el objeto de la imagen y clasificarla (si la clasificación es incorrecta, descartamos la imagen y pasamos a la próxima).

Experimentos y resultados

Para ver qué tan sensibles son los clasificadores frente a la pérdida de diferentes tipos de información, hicimos experimentos con 20 clases simplificadas de ImageNet, tomando 15 imágenes para cada clase. Tomamos cuatro modelos que usan redes neuronales profundas, que han logrado el mejor resultado sobre ImageNet en algún momento, y que han sido entrenados con las imágenes (completas) de entrenamiento de ImageNet. Los cuatro modelos, en orden de su rendimiento sobre ImageNet, son SqueezeNet, GoogLeNet, ResNet50, y SeNetResNet50. Se pueden ver ejemplos de las imágenes mínimas de cada modelo en la Figura 1 considerando varias formas de reducción.

Luego medimos la proporción de reducción para las imágenes mínimas positivas como el cociente entre el tamaño de la imagen original y la imagen mínima positiva (ambas comprimidas con PNG). Un menor cociente significa que el modelo es más robusto a la pérdida de información correspondiente. En la Figura 2, podemos ver los resultados, presentados como un diagrama de caja. Se puede ver que los humanos son mejores para clasificar imágenes con menos colores y resolución, pero que las máquinas pueden clasificar las imágenes basado en zonas más pequeñas. Estos resultados apoyan la observación de Geirhos *et al.* [4] de que la textura de la imagen es una característica importante para las redes neuronales profundas, las cuales pueden diferenciar, por ejemplo, entre el pelo de un gato y un perro. Por eso sólo necesitan una zona pequeña de una imagen, pero sufren más con una pérdida de resolución o color. Otra observación es que los modelos más robustos frente a la pérdida de información también tienen mejor rendimiento para las imágenes completas.

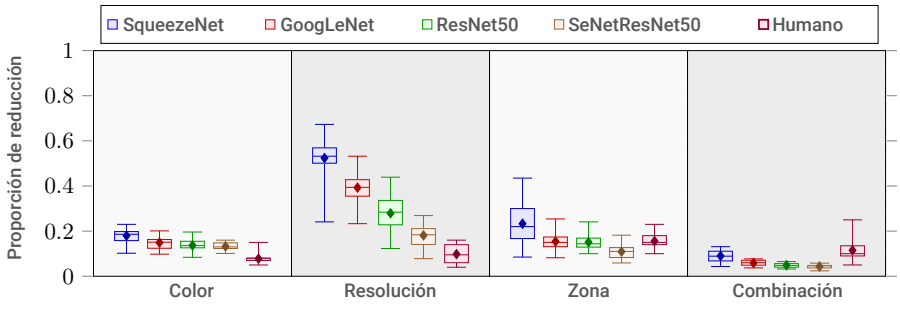


Figura 2. Proporción de reducción para las imágenes mínimas positivas.

Finalmente, hicimos un experimento usando cada clasificador para clasificar las imágenes mínimas positivas de los otros clasificadores. Se pueden encontrar los resultados completos en nuestro artículo [3]. En resumen, observamos que los humanos pueden clasificar mejor las imágenes mínimas positivas de las máquinas que al revés, logrando una precisión de 0,89-0,92 para color, 0,86-0,93 para resolución, 0,76-0,87 para zona, y 0,74-0,85 para combinación, con mejor precisión para las imágenes mínimas positivas, respectivamente, de SqueezeNet (más fáciles), GoogLeNet, ResNet50, y SeNetResNet50 (más difíciles). Al revés, clasificando las imágenes mínimas positivas de los humanos, los modelos de máquina lograron una precisión de 0,14-0,42 para color, 0,03-0,29 para resolución, 0,11-0,42 para zona, y 0,07-0,35 para combinación; los mejores modelos fueron, respectivamente, SeNetResNet50 (mayor precisión), ResNet50, GoogLeNet y SqueezeNet (menor precisión).

Conclusiones

¿Puede una máquina ver mejor que un humano? Es una pregunta cada vez más compleja, que puede ser interpretada de varias formas. En la Clasificación de Imágenes, nuestros resultados han indicado que los humanos proveen resultados más robustos frente a la pérdida de información. En la práctica, esto implica que los resultados dados por las redes neuronales profundas entrenadas y evaluadas en el contexto de conjuntos de imágenes completas pueden no aplicarse a condiciones reales, en las cuales un objeto (por ejemplo, una cara) está parcialmente oculto, o está a distancia, o iluminado parcialmente, etc.

Una pregunta que nos interesa ahora, entonces, es la siguiente: ¿se puede mejorar la robustez de los clasificadores de máquinas frente a la pérdida de información? Los modelos que usamos en este trabajo fueron entrenados sobre imágenes completas. Quizás se puedan entrenar las redes con imágenes reducidas o mínimas, para mejorar su robustez en situaciones de información parcial. ■

REFERENCIAS

[1] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael S. Bernstein, Alexander C. Berg, y Fei-Fei Li. 2015. ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision 115, 3 (2015), 211–252.

[2] Jie Hu, Li Shen, Samuel Albanie, Gang Sun, y Enhua Wu. 2019. Squeeze-andExcitation Networks. arXiV:1709.01507v4.

[3] Javier Carrasco, Aidan Hogan y Jorge Pérez. 2020. Laconic Image Classification: Human vs. Machine Performance. En el acta de la International Conference on Information and Knowledge Management (CIKM), Galway, Ireland, [Online], October 19–23, 2020.

[4] Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A. Wichmann, y Wieland Brendel. 2019. ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. En el acta de la International Conference on Learning Representations (ICLR). OpenReview.net.

Procesamiento de Lenguaje Natural: dónde estamos y qué estamos haciendo

FELIPE BRAVO-MÁRQUEZ Profesor Asistente del Departamento de Ciencias de la Computación de la Universidad de Chile e Investigador Joven del Instituto Milenio Fundamentos de los Datos.

JOCELYN DUNSTAN Profesora Asistente de la Iniciativa de Datos e Inteligencia Artificial de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile e Investigadora del Centro de Modelamiento Matemático.

El Procesamiento de Lenguaje Natural (PLN) es una rama de la Inteligencia Artificial (IA) centrada en el diseño de métodos y algoritmos que toman como entrada o producen como salida datos en la forma de lenguaje humano [1]. Esto puede venir en forma de texto o audio, y una vez que el audio es transcrito, ambos tipos de datos tienen un análisis común.

Tal como argumentan Julia Hirschberg y Chris Manning [2], tareas actuales donde el PLN entra en nuestras vidas son la traducción automática, los sistemas de pregunta-respuesta y la minería de texto en redes sociales. Ahondemos en la primera de ellas: la Web está en su mayoría en inglés, y el poder traducir páginas en forma casi instantánea es algo extraordinario. Traducir un texto no es fácil pues no hay una biyección entre palabras en ambos lenguajes, sino que una frase puede requerir menos palabras en un idioma que en otro (pensar por ejemplo traducir del español al inglés). Pero además, la traducción de una palabra requiere información del contexto en la que aparece para saber el sentido en la que se está usando. Asimismo, puede ocurrir que la palabra no tenga sentido en sí misma sino que en conjunto con la palabra que la acompaña (piense en las *phrasal verbs* del inglés). Actualmente los traductores automáticos usados por Google o DeepL están basados en sofisticadas redes neuronales.

PLN suele confundirse con otra disciplina hermana llamada Lingüística Computacional (LC). Si bien ambas están estrechamente relacionadas, tienen un foco distinto. La LC busca responder preguntas fundamentales sobre el lenguaje mediante el uso de la computación, es decir, cómo entendemos el lenguaje, cómo producimos lenguaje o cómo aprendemos lenguaje. Mientras que en PLN el foco está en resolver

problemas específicos, tales como las transcripción automática del habla, la traducción automática, la extracción de información de documentos y el análisis de opiniones en redes sociales. Es importante señalar que en PLN, el éxito de una solución se mide en base a métricas concretas (por ejemplo: qué tan similar es la traducción automática a una hecha por un humano) independientemente de si el modelo hace uso de alguna teoría lingüística.

Comprender y producir el lenguaje computacionalmente es extremadamente complejo. La tecnología más exitosa actualmente para abordar PLN es el aprendizaje automático supervisado que consiste en una familia de algoritmos que "aprenden" a construir la respuesta del problema en cuestión en base a encontrar patrones en datos de entrenamiento etiquetados.[1] Por ejemplo, si queremos tener un modelo que nos diga si un *tweet* tiene un sentimiento positivo o negativo respecto a un producto, primero necesitamos etiquetar manualmente un conjunto de *tweets* con su sentimiento asociado. Luego debemos entrenar un algoritmo de aprendizaje sobre estos datos para poder predecir de manera automática el sentimiento asociado a *tweets* desconocidos. Como se podrán imaginar, el etiquetado de datos es una parte fundamental de la solución y puede ser un proceso muy costoso, especialmente cuando se requiere conocimiento especializado para definir la etiqueta.

Los orígenes de PLN se remontan a los años cincuenta con el famoso test de Alan Turing: una máquina será considerada inteligente cuando sea capaz de conversar con una persona sin que ésta pueda determinar si está hablando con una máquina o un ser humano. A lo largo de su historia la disciplina ha tenido tres grandes periodos: 1) el racionalismo, 2) el empirismo, y 3) el aprendizaje profundo [3] que describimos a continuación.

El racionalismo abarca desde 1950 a 1990, donde las soluciones consistían en diseñar reglas manuales para incorporar mecanismos de conocimiento y razonamiento. Un ejemplo emblemático es el agente de conversación (o *chatbot*) ELIZA desarrollado por Joseph Weizenbaum que simulaba un psicoterapeuta rogeriano. Luego, a partir de la década de los noventa, el diseño de métodos estadísticos y de aprendizaje automático construidos sobre corpus llevan a PLN hacia un enfoque empirista. Las reglas ya no se construyen sino que se "aprenden" a partir de datos etiquetados. Algunos modelos representativos de esta época son los filtros de *spam* basados en modelos lineales, las cadenas de Markov ocultas para la extracción de categorías sintácticas y los modelos probabilísticos de IBM para la traducción automática. Estos modelos se caracterizaban por ser poco profundos en su estructura de parámetros y por depender de características manualmente diseñadas para representar la entrada.[2]

A partir del año 2010, las redes neuronales artificiales, que son una familia de modelos de aprendizaje automático, comienzan a mostrar resultados muy superiores en varias tareas emblemáticas de PLN [4]. La idea de estos modelos es representar la entrada (el texto) con una jerarquía de parámetros (o capas) que permiten encontrar representaciones idóneas para la tarea en cuestión, proceso al cual se refiere como "aprendizaje profundo". Estos modelos se caracterizan por tener muchos más parámetros que los modelos anteriores (superando la barrera del millón en algunos casos) y requerir grandes volúmenes de datos para su entrenamiento. Una gracia de estos modelos es que pueden ser preentrenados con texto no etiquetado como libros, Wikipedia, texto de redes sociales y de la Web para encontrar representaciones iniciales de palabras y oraciones (a lo que conocemos como *word embeddings*), las cuales pueden ser posteriormente adaptadas para la tarea objetivo donde sí se tienen datos etiquetados (proceso conocido como *transfer learning*). Aquí destacamos modelos como Word2Vec [5], BERT [6] y GPT-3 [7].

Este tipo de modelos ha ido perfeccionándose en los últimos años, llegando a obtener resultados cada vez mejores para casi todos los problemas del área [8]. Sin embargo, este progreso no ha sido libre de controversias. El aumento exponencial en la cantidad de parámetros[3] de cada nuevo modelo respecto a su predecesor, hace que los recursos computacionales y energéticos necesarios para construirlos sólo estén al alcance de unos pocos. Además, varios estudios han mostrado que estos modelos aprenden y reproducen los sesgos y prejuicios (por ejemplo: género, religión, racial) presentes en los textos a partir de los cuales se entrenan. Sin ir más lejos, la investigadora Timmnit Gebru fue despedida de Google cuando se le negó el permiso para publicar un artículo que ponía de manifiesto estos problemas [9].

1 | En PLN se le suele llamar a estos conjuntos de datos textuales (etiquetados o no etiquetados) como "corpus".

2 | La mayor parte de algoritmos de aprendizaje operan sobre vectores numéricos, donde cada columna es una característica del objeto a modelar. En PLN esas características pueden ser las palabras de una oración, las frases u otra propiedad (por ejemplo: el número de palabras con mayúsculas, la cantidad de emojis en un *tweet*, etc.).

3 | Word2Vec [5] tiene del orden de cientos de parámetros, BERT [6] tiene 335 millones de parámetros y GPT-3 [7] tiene 175 mil millones de parámetros.

Representations for Learning and Language (ReLeLa)[4] es un grupo de investigación del Departamento de Ciencias de la Computación (DCC) de la Universidad de Chile, donde también participan académicos y estudiantes de otros departamentos y centros. Sus miembros investigan varios temas en PLN: análisis de sentimiento y emociones en redes sociales, texto clínico, educación, textos legales, lenguas indígenas y el análisis de argumentos políticos.

Una línea de ReLeLa liderada por Jorge Pérez, ha sido el desarrollo de modelos preentrenados para el idioma español. Una contribución destacada ha sido BETO[5], la versión en español de BERT, que es ampliamente utilizado por investigadores y desarrolladores del mundo hispano.

En el ámbito del texto clínico, la creación de recursos para la extracción de información relevante requiere un trabajo fuertemente interdisciplinario. Recientemente fue presentado en el *workshop clínico de EMNLP*[6] el primer corpus clínico chileno etiquetado y resultados preliminares para el reconocimiento automático de entidades nombradas.

Finalmente, *The Word Embeddings Fairness Evaluation Framework* (WEFE)[7], es una herramienta de código abierto que permite medir y mitigar el sesgo de los modelos preentrenados señalados anteriormente. La principal característica de WEFE es estandarizar los esfuerzos existentes en un marco común para ser libremente utilizado.

A pesar de los grandes avances en los últimos años, aún estamos lejos de responder todas las interrogantes de PLN. En problemas como el diseño de *chatbots* las soluciones del estado del arte aún distan mucho de lo esperado y ni siquiera es claro cómo evaluarlas correctamente, luego para muchos otros problemas del mundo real simplemente no es posible obtener los recursos necesarios (datos etiquetados, hardware) para construir una solución adecuada. En RELELA confluyen visiones provenientes de la computación, las matemáticas, la lingüística y la salud para discutir esas interrogantes y sobre todo para mantenernos al día con los constantes avances del área. Todo esto ocurre en nuestros seminarios semanales donde escuchamos exposiciones de miembros del grupo o de algún charlista invitado. ■

REFERENCIAS

[1] Eisenstein, J. (2018). Natural language processing.

[2] Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. Science, 349(6245), 261–266.

[3] Deng, L., & Liu, Y. (Eds.). (2018). *Deep learning in natural language processing*. Springer.

[4] Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., and Kuksa, P. (2011). Natural language processing (almost) from scratch. Journal of machine learning research, 12(Aug):2493–2537.

[5] Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2 (NIPS'13).

[6] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 4171–4186.

[7] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., et al. (2020). Language models are few-shot learners. In Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 202.

[8] NLP-progress: Repository to track the progress in Natural Language Processing (NLP), including the datasets and the current state-of-the-art for the most common NLP tasks: http://nlpprogress.com/.

[9] Bender, Emily M., et al. (2021). "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜." *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*.

4 | https://relela.com/.

5 | https://github.com/dccuchile/beto.

6 | https://www.aclweb.org/anthology/2020.clinicalnlp-1.32/.

7 | https://wefe.readthedocs.io/en/latest/.

Inteligencia artificial para restauración de material arqueológico

ALEXIS MENDOZA Estudiante de pregrado de la Escuela de Ciencia de la Computación, Universidad Nacional San Agustín, Perú.
ALEXANDER APAZA Estudiante de pregrado de la Escuela de Ciencia de la Computación, Universidad Nacional San Agustín, Perú.
IVÁN SIPIRÁN Profesor Asistente del Departamento de Ciencias de la Computación, Universidad de Chile.
CRISTIÁN LÓPEZ Profesor Asistente del Departamento de Ingeniería, Universidad de Ingeniería y Tecnología, Perú.

En 2018, el museo Josefina Ramos de Cox en Lima - Perú inició un proceso de digitalización de los objetos arqueológicos que albergan en su colección. El museo administra más de siete mil piezas provenientes de diferentes culturas prehispánicas, principalmente culturas de la costa central del Perú. Para el proceso de digitalización, el museo usó un escáner 3D de escritorio que utiliza tecnología de luz estructurada. Sin embargo, el proceso de digitalización no se desarrolló de forma satisfactoria por dos razones:

1. La mayoría de los objetos eran frágiles y, al no poder sostenerse sobre la base del escáner, se tuvo que colocar bases artificiales. Estas bases artificiales fueron posteriormente removidas en las superficies 3D generadas, dejando grandes porciones de la base de los objetos sin información.

2. El escáner de luz estructurada tiene problemas para escanear superficies cuyo ángulo con respecto al haz de luz es casi perpendicular. Por lo tanto, hay bases de objetos que no fueron correctamente escaneadas por la limitación del escáner.

El problema en la digitalización trajo como consecuencia que un gran número de objetos tengan una superficie incompleta después del escaneo (ver Figura 1). Nosotros propusimos una forma de solucionar el problema de la geometría faltante desde un enfoque basado en datos y usando inteligencia artificial.

Nuestra propuesta

Nuestro método consiste de una red neuronal que recibe un objeto 3D con superficie incompleta y produce el objeto completo reparado. Nuestra premisa es que si contamos con suficientes ejemplos de objetos dañados y objetos completos, la red neuronal puede encontrar una buena correspondencia entre la geometría de la superficie incompleta y

Figura 1. Vista frontal y superior de algunos objetos escaneados. Note la falta de geometría en la base de los objetos.

la superficie de los objetos completos. Además, si seguimos un protocolo de entrenamiento adecuado, podemos esperar que la red neuronal generalice bien a diferentes geometrías faltantes.

El problema es que la colección escaneada del museo Josefina Ramos de Cox no contiene muchos ejemplos de objetos completos, como para permitir hacer un entrenamiento adecuado de una red neuronal. En este punto, hicimos una observación clave para solucionar el problema. Lo que requerimos de la red neuronal es que aprenda la estructura de objetos arqueológicos, por lo que cualquier otro conjunto de datos con estructura similar podría servir para nuestro cometido. Así, logramos recolectar un conjunto de 1458 objetos desde el 3D Pottery Benchmark [1] y las clases "Bowl" y "Jar" del dataset ShapeNet [2]. Todos estos objetos tienen estructura común a objetos arqueológicos y sirvieron para entrenar nuestra red neuronal.

Con respecto a la arquitectura de la red neuronal, típicamente el problema de "*shape completion*" se aborda desde una perspectiva de un modelo tipo *encoder-decoder*, en donde el encoder procesa la geometría de entrada y la transforma en un vector numérico. Posteriormente, el vector numérico es la entrada al decoder, que finalmente reconstruye la geometría completa [3, 4]. Sin embargo, un problema con este tipo de arquitectura es que generan una representación transformada de la geometría completa. En nuestro caso, la geometría de entrada no tiene que ser cambiada ni transformada, y más bien lo que necesitamos es generar una buena representación de la superficie que falta. Es así que nosotros presentamos una nueva arquitectura para este problema específico, en donde una primera red neuronal produce una región faltante candidata. La unión del objeto incompleto y la región candidata es posteriormente refinada con una segunda red neuronal, la cual produce el objeto completo. Ambas redes neuronales son entrenadas en conjunto y en forma *end-to-end*. Para la representación de los modelos 3D, escogimos las nubes de puntos [5]. La arquitectura puede verse en la Figura 2.

Para entrenar este modelo, usamos el conjunto de datos recolectado y

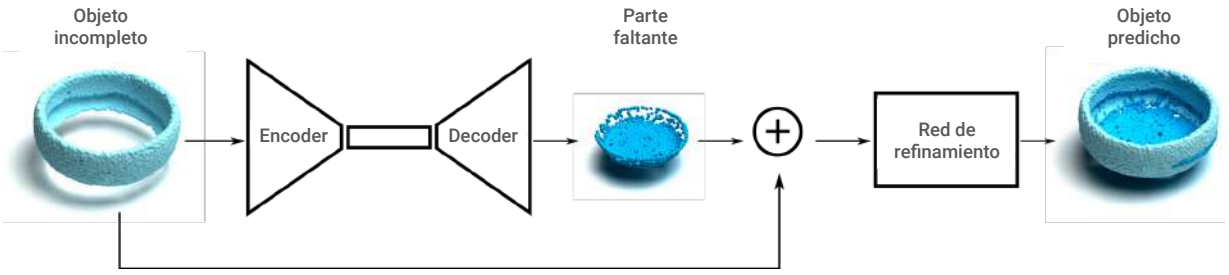


Figura 2. Arquitectura de nuestra red neuronal. El modelo consiste en un *encoder-decoder* para generar la parte faltante a partir del objeto incompleto. Ambos objetos son luego usados por la red de refinamiento para obtener el objeto reparado final.

Figura 3. Ejemplos de objetos reparados con nuestra herramienta.

realizamos la generación de pares de entrenamiento (objeto incompleto, objeto completo) durante el mismo entrenamiento. Creamos un protocolo para generar pares aleatorios de objetos, aplicando un algoritmo que simula la eliminación de geometría en la base de un objeto de entrada. Este algoritmo nunca genera dos objetos iguales durante el entrenamiento, por lo que esto garantiza que la red no memorice los ejemplos de entrenamiento.

Una vez que la red fue entrenada, usamos el conjunto de objetos arqueológicos del museo como objetos de prueba. Como la red procesa nubes de puntos, implementamos un algoritmo que reconstruye la superficie de los objetos 3D. La Figura 3 muestra algunos resultados de nuestro método.

Consideraciones finales

Abordamos un problema de restauración de piezas arqueológicas desde una perspectiva de datos. Este trabajo se pudo llevar a cabo gracias a los recientes avances en análisis de formas y procesamiento geométrico a través del uso de técnicas de aprendizaje automático. Nuestros resultados muestran que las redes neuronales que procesan geometría pueden extraer información de estructura de los objetos. Esta estructura puede ser empleada para el diseño asistido por computadora, y específicamente en nuestro caso fue útil para predecir la geometría faltante de objetos con defectos de escaneo. ■

REFERENCIAS

[1] Koutsoudis A., Pavlidis G., Liami V., Tsiafakis D., Chamzas C., "*3D Pottery content-based retrieval based on pose normalisation and segmentation*". Journal of Cultural Heritage, 11(3), pp 329-338, 2010.

[2] Chang A., Funkhouser T., Guibas L., Hanrahan P., Huang Q., Li Z., Savarese S., Savva M., Song S., Su H., Xiao J., Yi L., Yu F., "*ShapeNet: An Information-Rich 3D Model Repository*". CoRR abs/1512.03012. Arxiv, 2015.

[3] Yuan W., Khot T., Held D., Mertz C., Hebert M., "*PCN: Point Completion Network*". In Proc: *International Conference on 3D Vision (3DV)*, pp. 728-737. 2018.

[4] Tchapmi L., Kosaraju V., Rezatofighi H., Reid I., Savarese S., "*TopNet: Structural Point Cloud Decoder*". *In Proc: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 383-392. 2019.

[5] Qi R., Su H., Kaichun M., Guibas L., "*PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation*". *In Proc: IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 77-85. 2017.

Neuroevolución: ¿Cómo evitar los datos masivos de entrenamiento?

ALEXANDRE BERGEL Profesor Asociado del Departamento de Ciencias de la Computación de la Universidad de Chile.

Contexto

Según la teoría de Darwin, el cerebro de los mamíferos es el resultado de una larga evolución. Frente a cualquier otra especie, los humanos tienen el cerebro más grande en relación a su peso. Hace decenas de milenios, nuestro cerebro no tenía la sofisticación que tiene hoy. El cerebro evolucionó, en parte, para solucionar problemas complejos como la necesidad de los humanos de comunicarse en forma eficiente. Siguiendo un proceso de evolución similar al de nuestro cerebro, la neuroevolución es una técnica de la inteligencia artificial que combina un algoritmo genético con una red neuronal. Su idea central es producir modelos que sean lo suficientemente desarrollados para solucionar un problema que no se puede expresar a través de ejemplos. Es una idea casi opuesta a la forma en que se entrena un modelo con grandes cantidades de imágenes o de texto, como se hace en el área de *deep learning*.

Ejemplo y aplicaciones

Consideren la red neuronal de la Figura 1. Esta red describe el comportamiento del operador booleano *XOR*, usando una función de activación de tipo *step*. Tiene, además, nueve parámetros, tres por cada neurona. Un algoritmo de aprendizaje, como el *backpropagation* usado en *deep learning*, tendrá que deducir estos nueve parámetros desde un conjunto de ejemplos. En este caso, tener ejemplos no representa un problema para entrenar la red, pero en otros casos, según el problema a abordar, tener ejemplos puede representar un lujo que no siempre es alcanzable.

La neuroevolución es una técnica alternativa al *backpropagation* para deducir estos nueve parámetros y consiste en la aplicación de un algoritmo genético con redes neuronales. En vez de entrenar una red usando mecanismos de aprendizaje, la neuroevolución usa un algoritmo evolutivo para buscar los parámetros que generan redes de “mejor calidad”.

Un algoritmo genético es una metáfora computacional del mecanismo de evolución natural, tal como lo describió Charles Darwin. En la naturaleza, los individuos más fuertes tienen mayores probabilidades de sobrevivir y de reproducirse. Aplicado a nuestro ejemplo de

redes naturales, un individuo es una serie de nueve números y la probabilidad de evolucionar depende de la cantidad de errores que comete la red neuronal bajo este individuo.

En cada generación, los individuos más fuertes (i.e., las redes neuronales que cometen menos errores) se combinan usando operaciones genéticas, tales como la mutación y el *cross-over*. La población inicial de individuos está compuesta por series de nueve números aleatorios, pero en cada generación se genera una red mejor, que produce menos errores que en la generación anterior.

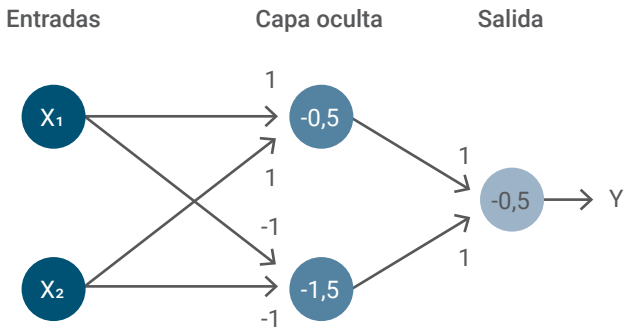


Figura 1. Red neuronal que simula al operador booleano *XOR*.

Algoritmos sofisticados de neuroevolución, como NEAT y HyperNEAT, permiten evolucionar no solamente los parámetros, sino también la topología de la red, algo que no se puede lograr con el *deep learning* clásico.

En el grupo ISCLab[1] del Departamento de Ciencias de la Computación (DCC) de la Universidad de Chile, usamos la neuroevolución para desarrollar inteligencia artificial de videojuego, estilo Mario Bros. La neuroevolución es particularmente conveniente para producir dicho tipo de IA ya que, en comparación al *deep learning*, no requiere datos de jugadas.

Por otro lado, estamos desarrollando técnicas de visualización que permiten caracterizar el proceso de evolución. La neuroevolución, como cualquier otro algoritmo de *machine learning*, es una caja negra que entrega un resultado sin dar cuenta del camino tomado para obtener dicho resultado. Nuestras visualizaciones ayudan a entender las diferentes decisiones tomadas por el algoritmo de neuroevolución, lo que ayuda a explicar su resultado.

Beneficios

La neuroevolución no tiene las limitaciones que imponen un uso de cantidades *masivas* de datos. Un modelo basado en neuroevolución puede superar a un modelo basado en ejemplos producidos por humanos. Ejemplos prominentes de esta situación son la robótica y los videojuegos. Si un jugador virtual tuviese que aprender de los humanos cómo jugar, no lograría superarlos. Pero un algoritmo evolutivo (al que la neuroevolución pertenece) puede superar, y por mucho, a los mejores jugadores del mundo. AlphaGo y Dota2 demuestran la amplia capacidad de los algoritmos evolutivos para superar a los humanos.

El artículo "Designing neural networks through neuroevolution", publicado en 2019 en la revista Nature Machine Intelligence, describe los últimos progresos en el área de la neuroevolución. Además de presentar una retrospectiva de cómo la naturaleza y la evolución del cerebro han tenido un enorme impacto en el área de la inteligencia artificial, este artículo describe una extraordinaria forma de acercarse a una inteligencia artificial genérica. Ahora, es reconocido que la neuroevolución es competidora de las técnicas modernas usadas en aprendizaje supervisado, al que pertenecen las técnicas de aprendizaje de redes neuronales. ■

1 | https://isclab.dcc.uchile.cl/.

Inteligencia artificial en la educación

JÉRÉMY BARBAY Profesor Asistente del Departamento de Ciencias de la Computación de la Universidad de Chile.

La tecnología siempre se ha incorporado a la docencia de manera desigual, y las técnicas de Inteligencia Artificial (IA) no son una excepción. Con el fin de contribuir a reducir dicha desigualdad, presentamos una vista superficial de: 1) algunas técnicas de inteligencia artificial, 2) algunos sistemas de manejo del aprendizaje, y 3) algunas aplicaciones de técnicas de IA a lo señalado en el punto 2. Con la finalidad de (intentar) guiar desarrollos futuros, presentamos una discusión corta sobre los desafíos presentes y futuros de las técnicas de inteligencia artificial sobre los sistemas de manejo de aprendizaje.

Historia y definiciones

Tecnologías educacionales

El campo de "tecnologías educacionales" corresponde al estudio y la práctica ética de facilitar la educación y mejorar el rendimiento creando, usando y manejando los recursos y procesos adecuados. Desde la perspectiva del uso de la tecnología en educación, tecnologías educacionales se puede entender como el uso de tecnologías existentes y emergentes para mejorar la experiencia de aprendizaje en una variedad de contextos instruccionales, como el aprendizaje formal, informal, no-formal, a demanda (*on-demand*) o *"just-in-time"* [1].

Tales tecnologías educacionales incluyen una gran variedad de dominios de desarrollo. Respecto al material, se consideran una gran cantidad de dispositivos, desde proyectores de apuntes copiados sobre láminas transparentes, computadoras personales e interconectadas, hasta tecnologías "inteligentes" como teléfonos, entornos virtuales, computación en la nube, aparatos *"wearable"* y *"location-aware"*. Respecto del software, se considera, por un lado el software dirigido a quienes aprenden, como los software de simulación y de visualización, y las interfaces de gamificación mejorando la motivación; y, por otro lado, el software dirigido a la administración del aprendizaje, con los "*Learning Management Systems*" (LMS)[1] y su integración vía "Learning Tools Interoperability" (LTI).[2]

1 | https://en.wikipedia.org/wiki/Learning_management_system.

2 | https://en.wikipedia.org/wiki/Learning_Tools_Interoperability.

La digitalización del material educativo, una tendencia que existía pero se desarrollaba relativamente lenta hasta 2019 [2], se ha acelerado con la transición súbita hacia la docencia online en el contexto de la pandemia por COVID-19. En este contexto, se han digitalizado muchos aspectos de la docencia. Por un lado, las charlas, tradicionalmente en anfiteatros y en vivo, y raramente grabadas, han sido reemplazadas en muchos casos por la difusión en tiempo real de tales charlas en video, y en otros casos por la difusión de cápsulas de videos cortas, grabadas y editadas con anticipación: en ambos casos, los alumnos pueden mirar tales videos en momentos de su elección, desde su hogar, y muchas veces las ven mientras hacen otras actividades y/o en modo acelerado. Por otro lado, las evaluaciones teóricas, tradicionalmente entregadas sobre papel, en instancias de exámenes presenciales, están siendo reemplazadas por entregas digitales, generando sospechas de copias y de usurpación de identidades en los cuerpos docentes.

En tal desarrollo de las tecnologías educacionales, era esperable ver llegar las técnicas de inteligencia artificial, las cuales intentamos definir en la siguiente sección, antes de desarrollar sus interacciones con el campo de *"Learning Management Systems"* en la sección "Aplicaciones de la IA a los LMS".

Técnicas de inteligencia artificial

Conviene primero aclarar el concepto de "inteligencia artificial". En la vida diaria, el término se aplica cuando una máquina imita las funciones "cognitivas" que los humanos asocian con mentes humanas, como por ejemplo: "percibir", "razonar", "aprender" y "resolver problemas" [4]. Una definición más formal y menos antropomórfica sería *"la capacidad de un sistema para interpretar correctamente datos externos, para aprender de dichos datos y emplear esos conocimientos para lograr tareas y metas concretas a través de la adaptación flexible"*. En ambos casos, mientras que las máquinas se vuelven más capaces, tecnologías que alguna vez se consideraban del campo de "inteligencia artificial" se reevalúan.

La expresión "inteligencia artificial" fue introducida en 1956 por John McCarthy, quien la definió como *"la ciencia e ingenio de hacer máquinas inteligentes, especialmente programas de cómputo inteligentes"*. Pero el concepto existía desde hace mucho más tiempo, lo que hace que siga evolucionando en paralelo con las tecnologías [3].

En 2021, los objetivos de "inteligencia artificial" se pueden clasificar en cuatro tipos [4]:

- *Sistemas que piensan como humanos.* Estos sistemas tratan de emular el pensamiento humano, por ejemplo, las redes neuronales artificiales. La automatización de actividades que vinculamos con procesos de pensamiento humano, actividades como la toma de decisiones, resolución de problemas y aprendizaje.

- *Sistemas que actúan como humanos.* Estos sistemas tratan de actuar como humanos, es decir, imitan el comportamiento humano, por ejemplo, la robótica. El estudio de cómo lograr que los computadores realicen tareas que, por el momento, los humanos hacen mejor.

- *Sistemas que piensan racionalmente.* Esto es, con lógica (idealmente), tratan de imitar el pensamiento racional del ser humano, por ejemplo, los sistemas expertos. El estudio de los cálculos que hacen posible percibir, razonar y actuar.

- *Sistemas que actúan racionalmente.* Tratan de emular de forma racional el comportamiento humano, por ejemplo los agentes inteligentes. Está relacionado con conductas inteligentes en artefactos.

En la siguiente sección veremos cómo las técnicas de inteligencia artificial se han relacionado y siguen relacionándose con las técnicas de educación y de aprendizaje.

Aplicaciones de la IA a los LMS

Desde muy temprano se relacionaron los temas de educación (humana) e inteligencia artificial, quizás porque en ambos casos se trata de desarrollar habilidades "inteligentes"', ya sea en humanos o en máquinas. Seymour Papert, uno de los cofundadores del Instituto de Inteligencia Artificial del MIT, en 1963 (con Marvin Minsky, considerado uno de los padres de la inteligencia artificial[3], había tenido previamente un rol mayor en la evaluación y el desarrollo de técnicas de educación, en colaboración con el psicólogo educativo Piaget.[4]

En 2021, técnicas de inteligencia artificial presentan aplicaciones en varios aspectos de la docencia. En un survey publicado en 2020, Chen et al. [5] describen varias aplicaciones de inteligencia artificial en áreas relacionadas con la educación, en particular aplicadas a los aspectos de la administración de la docencia. Tales aplicaciones permiten, entre otros, detectar ocurrencias de plagio, automatizar algunos aspectos de la evaluación de trabajos, e identificar a un alumno presente cuyo perfil sea similar al perfil de alumnos anteriores que tuvieron problemas en fases siguientes.

Por otro lado, software como Duolingo[5] usa técnicas de gamificación para mantener la motivación de sus alumnos, y técnicas de repetición espaciada [6] para programar qué ejercicio darle a un alumno en función de modelos.

En el futuro, técnicas de inteligencia artificial tendrán otras aplicaciones en educación. Investigadores como la Dra. Shaghayegh Sahebi están proponiendo diseñar, desarrollar y evaluar sistemas capaces de realizar recomendaciones personalizadas de material docente en función de varios parámetros [7].

Conclusiones

Las técnicas descritas como "inteligencia artificial" no son más que nuevas tecnologías que apuntan a acercar las capacidades de las máquinas a las capacidades de los humanos. En varias épocas se sobrepromotió lo que se podía lograr con dichas técnicas, y la época presente no es una excepción. Pero aún permiten automatizar algunas tareas humanas, y apoyar otras.

El área de la educación, y en particular el área de la educación en línea, tiene un gran potencial de mejoras vía técnicas digitales en general, y técnicas propias de "inteligencia artificial" en particular, y ha sido un poco lenta en adoptar dichas técnicas. Es esperable que con la digitalización acelerada debido a la pandemia por COVID-19, dicha transición se vea acelerada.

Como siempre con la tecnología, será importante no dejar el efecto de novedad, ni quitar el foco de problemas importantes existentes (por ejemplo, desigualdades) ignorados o amplificados por nuevas técnicas, ni de nuevos problemas creados por dichas técnicas (por ejemplo, sesgos en favor de minorías producidos por técnicas de inferencias, impacto ecológico de las digitalizaciones, etc.). ■

REFERENCIAS

[1] R. Huang, J. Spector y J. Yang (2019). Educational Technology: A Primer for the 21st Century. 10.1007/978-981-13-6643-7.

[2] J. Barbay y V. Peña-Araya (2019). El Académico Digital. En Revista Bits de Ciencia nº18.

[3] Historia de la inteligencia artificial. En *Wikipedia*. Accedido desde https://es.wikipedia.org/wiki/Historia_de_la_inteligencia_artificial, [2021-04-19 Mon].

[4] Inteligencia artificial. En *Wikipedia*. Accedido desde https://es.wikipedia.org/wiki/Inteligencia_artificial, last accessed, [2021-04-19 Mon].

[5] L. Chen, P. Chen y Z. Lin. (2020). Artificial Intelligence in Education: A Review. En IEEE Access, vol. 8, pp. 75264–75278, 10.1109/ACCESS.2020.2988510.

[6] Spaced repetition. En *Wikipedia*. Accedido desde https://en.wikipedia.org/wiki/Spaced_repetition, [2021-04-19 Mon].

[7] https://www.nsf.gov/awardsearch/showAward?AWD_ID=2047500, [2021-04-19 Mon].

3 | https://es.wikipedia.org/wiki/Marvin_Minsky.

4 | https://es.wikipedia.org/wiki/Seymour_Papert.

5 | https://www.duolingo.com/.

Aprendizaje de representaciones en grafos y su importancia en el análisis de redes

MARCELO MENDOZA Profesor Asociado del Departamento de Informática de la Universidad Técnica Federico Santa María e Investigador Asociado del Instituto Milenio Fundamentos de los Datos.

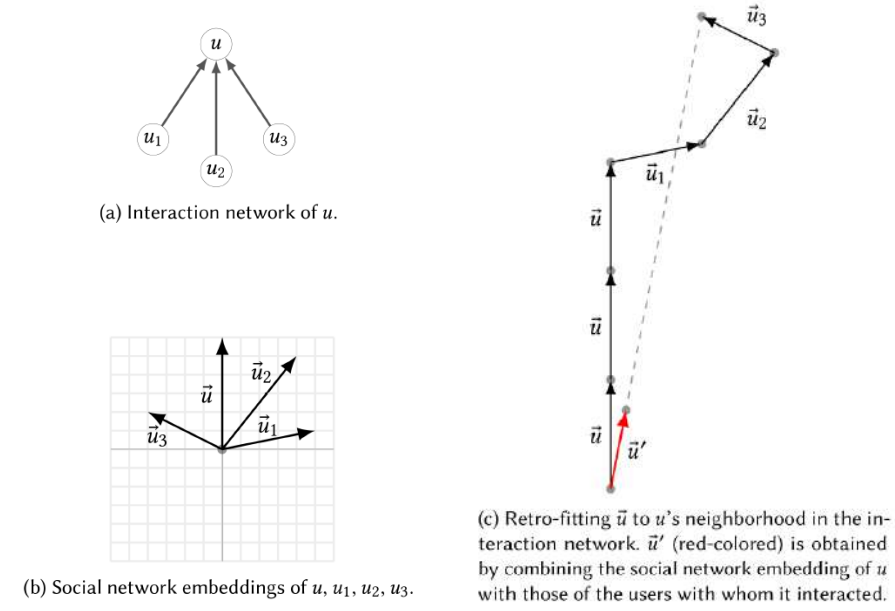
Una de las líneas de investigación en inteligencia artificial más fructíferas de la última década es el aprendizaje de representaciones. Mostraremos dos ejemplos en los cuales el aprendizaje de representaciones de nodos en grafos ha permitido abordar exitosamente tareas de análisis de redes.

Detección de *bots*

Los *bots* tienen un nefasto efecto en la diseminación de información engañosa o tendenciosa en redes sociales [1]. Su objetivo es amplificar la alcanzabilidad de campañas, transformando artificialmente mensajes en tendencias. Para ello, las cuentas que dan soporte a campañas se hacen seguir por cuentas manejadas por algoritmos. Muchas de las cuentas que siguen a personajes de alta connotación pública son *bots*, las cuales entregan soporte a sus mensajes con *likes* y *retweets*. Cuando estos mensajes muestran un inusitado nivel de reacciones, se transforman en tendencias, lo cual aumenta aún más su visibilidad. Al transformarse en tendencias, su influencia en la red crece, produciendo un fenómeno de bola de nieve.

La detección de *bots* ha sido una tarea difícil. Mientras que las primeras generaciones de *bots* eran sencillas de detectar, las nuevas generaciones de *bots*, conocidas como *social bots*, alternan periodos de propaganda y periodos de baja actividad [2]. En estos últimos, los *bots* muestran un comportamiento cercano al de un usuario promedio, con participación esporádica en la red. En periodos de campaña, la actividad de estas cuentas aumenta. El cambio en el régimen de interacciones es una pista que nosotros usamos para detectarlos.

En [3], mostramos cómo extender una representación de nodos aprendida a partir de la red de conexiones sociales en Twitter. La estrategia de aprendizaje usada se denomina ComplEx [4], la cual permite aprender *node embeddings* de la red de conexiones para predicción de *links*. Para capturar el régimen de interacciones entre cuentas, extendemos ComplEx reescalando los *node embeddings* en la dirección de los vecinos con los cuales tienen más interacciones. La Figura 1 muestra la estrategia de reescalamiento basada en interacciones, lo cual permite recalcular los *node embeddings* combinando ambas redes (social e interacción). Para aprender los *node embeddings* usamos una estrategia denominada *retrofitting* [5], que busca una

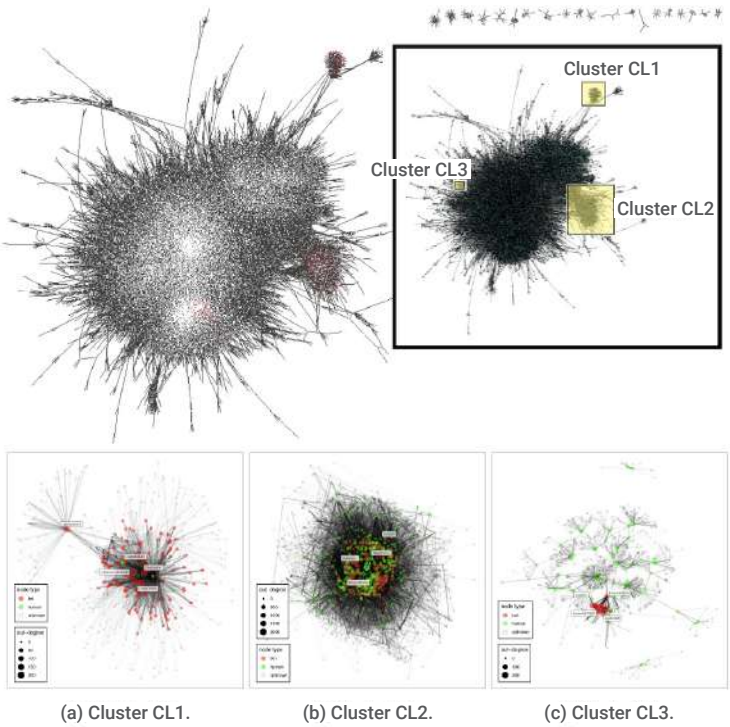


(a) Interaction network of u.

(b) Social network embeddings of u, u_1, u_2, u_3.

(c) Retro-fitting $\vec{u}$ to u's neighborhood in the interaction network. $\vec{u}'$ (red-colored) is obtained by combining the social network embedding of u with those of the users with whom it interacted.

Figura 1. Extensión de ComplEx [4] que incorpora la red de interacciones entre usuarios de Twitter.



(a) Cluster CL1. (b) Cluster CL2. (c) Cluster CL3.

Figura 2. Red de proximidad entre *node embeddings* en Twitter, que muestra tres *clusters* con presencia de *bots* (nodos rojos). Mientras que el *cluste*r 1 (CL1) no logra interactuar con humanos (nodos verdes), los *clusters* 2 (CL2) y 3 (CL3) se mimetizan, promoviendo contenido propagandístico.

representación consistente entre ambas fuentes de información.

Para detectar *bots*, aplicamos un algoritmo de propagación de etiquetas en la red de proximidad de *node embeddings*. El método de propagación permite trabajar con un número reducido de nodos etiquetados como *bots*, usando una estrategia semisupervisada sobre la red. La estrategia semisupervisada permite que el método funcione sobre redes de enorme tamaño con sólo una fracción de sus nodos etiquetados por expertos (app. 1% del total de la red). Mostramos que el método de imputación de etiquetas es análogo a una estrategia de paso de mensajes en una red neuronal de grafos que aborda una tarea de clasificación de nodos [6].

Nuestro método superó al estado del arte (Botometer [7] y Holoscope [8]). Su principal habilidad está en la detección de *botnets*, lo cual le permite sacar ventaja de sus más directos competidores que abordan la tarea como clasificación de nodos. El método de propagación de etiquetas tiene la ventaja de identificar grupos de cuentas *clusterizadas* según interacciones inusuales, detectando patrones de coordinación temporal. La Figura 2 muestra una red de proximidad entre *node embeddings* y tres *clusters* con alta presencia de *bots* (nodos rojos) en Twitter. Mientras que el *cluster* 1 (CL1) es una *botnet* que no ha logrado interactuar con humanos (nodos verdes), los *clusters* 2 (CL2) y 3 (CL3) muestran una mimetización de los *bots* en las redes de humanos, con interacción cruzada entre ambos tipos de usuarios.

Predictibilidad en redes sociales *offline*

En [9], analizamos las relaciones filiales entre personas, observables a través de los vínculos de apellidos paternos-ma-

ternos. La red construida con los datos del servicio electoral y cruzada con datos del Indice de Bienestar Territorial nos permitió construir un mapa de las conexiones familiares de los habitantes de la Región Metropolitana. Usando el método de Mateos *et al.* [10], identificamos los vínculos cuyas ocurrencias superaban el valor esperado dado por una red de conexiones aleatorias. Una vez construida la red, visualizamos su estructura agrupando nodos según modularidad. Las comunidades detectadas muestran etnias y también una fuerte *clusterización* de apellidos de clase alta según índice socioeconómico (ver Figura 3, al tope).

La misma red, ahora *clusterizada* según ingreso socioeconómico (ver Figura 3, al medio), muestra dos particiones, una con una fuerte interacción entre apellidos poco frecuentes y muchos nodos articuladores (comunidad azul de los tres deciles de ingreso más alto), y una partición mucho más desarticulada, con una vinculación más débil entre apellidos y menos nodos articuladores (comunidad roja de los siete deciles más bajos de ingreso). Estudiamos la predictibilidad de esta red, donde la tarea corresponde a predecir vínculos entre familias no conectadas (*link prediction*). Para hacer esto, aplicamos una técnica de aprendizaje de representaciones de nodos basada en factorización tensorial denominada método de TuckER [11]. Probamos el desempeño de otros métodos de representación a nivel de nodos, como ComplEx [4], RESCAL [12] y RotatE [13], usados en *knowledge-base completion*. TuckER mostró mejor desempeño en *link prediction* que sus competidores, factor atribuible a su habilidad de trabajar con datos *sparse*.

Al pie de la Figura 3 mostramos los resultados de predicción de vínculos segmentados por decil de ingreso. Los deciles de mayor ingreso (d1 - d3) muestran mejor predictibilidad, la cual disminuye progresivamente para los deciles de menor ingreso (d4 – d10).

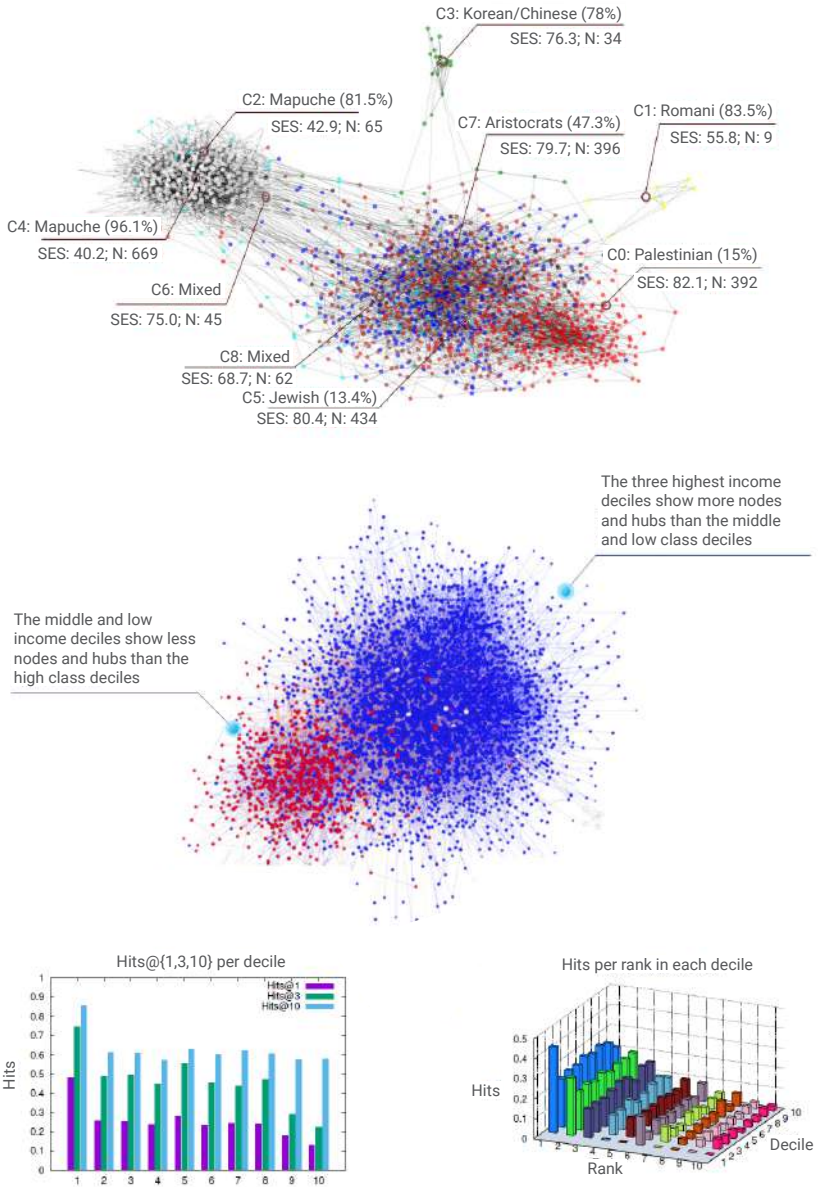


Figura 3. Redes de vínculos paternos-maternos en la Región Metropolitana (al tope), la misma red *clusterizada* según ingreso socioeconómico (al medio), y la predictibilidad de vínculos usando TuckeER [11] (al pie).

Conclusión

La inteligencia artificial a través de su área denominada aprendizaje de representaciones ofrece enormes posibilidades en tareas complejas, tanto en redes sociales en línea como en redes *offline*. Su habilidad para codificar características esenciales en distintos dominios permite generar representaciones que mejoran las posibilidades de análisis de datos. ■

REFERENCIAS

[1] Stefano Cresci: A decade of social bot detection. Commun. ACM 63(10): 72–83 (2020).

[2] Stefano Cresci, Roberto Di Pietro, Marinella Petrocchi, Angelo Spognardi, Maurizio Tesconi: The Paradigm-Shift of Social Spambots: Evidence, Theories, and Tools for the Arms Race. WWW (Companion Volume) 2017: 963–972.

[3] Marcelo Mendoza, Maurizio Tesconi, Stefano Cresci: Bots in Social and Interaction Networks: Detection and Impact Estimation. ACM Trans. Inf. Syst. 39(1): 5:1–5:32 (2020).

[4] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, Guillaume Bouchard: Complex Embeddings for Simple Link Prediction. ICML 2016: 2071–2080.

[5] Manaal Faruqui, Jesse Dodge, Sujay Kumar Jauhar, Chris Dyer, Eduard H. Hovy, Noah A. Smith: Retrofitting Word Vectors to Semantic Lexicons. HLT-NAACL 2015: 1606–1615.

[6] Franco Scarselli, Sweah Liang Yong, Marco Gori, Markus Hagenbuchner, Ah Chung Tsoi, Marco Maggini: Graph Neural Networks for Ranking Web Pages. Web Intelligence 2005: 666–672.

[7] Onur Varol, Emilio Ferrara, Clayton A. Davis, Filippo Menczer, Alessandro Flammini: Online Human-Bot Interactions: Detection, Estimation, and Characterization. ICWSM 2017: 280–289.

[8] Shenghua Liu, Bryan Hooi, Christos Faloutsos: HoloScope: Topology-and-Spike Aware Fraud Detection. CIKM 2017: 1539–1548.

[9] Naim Bro, Marcelo Mendoza. Surname affinity in Santiago, Chile: A network-based approach that uncovers urban segregation. PLOS ONE, 16(1): e0244372, 2021.

[10] Pablo Mateos, Paul Longley, David O'Sullivan. Ethnicity and population structure in personal naming networks. PLOS ONE, 6(9): e22943, 2011.

[11] Ivana Balazevic, Carl Allen, Timothy M. Hospedales: TuckER: Tensor Factorization for Knowledge Graph Completion. EMNLP/IJCNLP (1) 2019: 5184–5193.

[12] Maximilian Nickel, Volker Tresp, Hans-Peter Kriegel: A Three-Way Model for Collective Learning on Multi-Relational Data. ICML 2011: 809–816

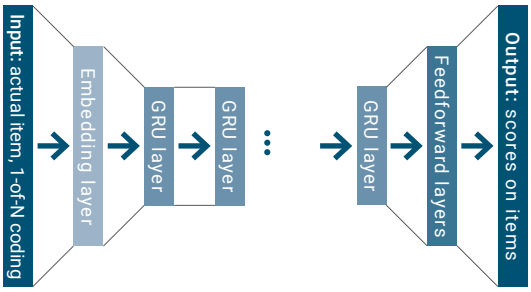
[13] Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, Jian Tang: RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space. ICLR 2019.

Aprendizaje profundo en sistemas de recomendación

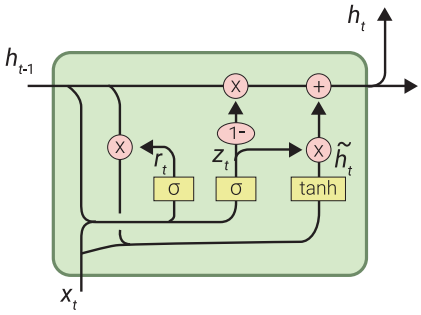
DENIS PARRA Profesor Asociado del Departamento de Ciencia de la Computación de la Pontificia Universidad Católica de Chile e Investigador Adjunto del Instituto Milenio Fundamentos de los Datos.

Corría el año 2010 y yo cursaba mi doctorado enfocado en personalización y sistemas de recomendación en la Universidad de Pittsburgh, ubicada en la ciudad homónima (Pittsburgh) al oeste del estado de Pennsylvania en Estados Unidos. Las técnicas más avanzadas de mi tema de investigación eran del área conocida como Aprendizaje Automático (en inglés, *Machine Learning*), por lo que sentía la necesidad de tomar un curso avanzado para completar mi formación. En el semestre de otoño finalmente me inscribí en el curso de Aprendizaje Automático, y gracias a un convenio académico pude cursarlo en la universidad vecina, Carnegie Mellon University. Yo estaba realmente emocionado de tomar un curso en un tema de tan creciente relevancia en unas de las mejores universidades del mundo en el área de computación. Recuerdo que vimos muchas técnicas que permitían aprender modelos a partir de datos, con especial énfasis en modelos gráficos —por ejemplo, el famoso *Latent Dirichlet Allocation* [1]— así como en métodos *kernel* como *Support Vector Machines* (SVM). Casi al final del curso, tuvimos una clase algo tímida sobre redes neuronales artificiales, un método interesante pero que poca gente usaba. Las redes neuronales artificiales datan de los años cincuenta [2], renacieron en los ochenta luego del invierno de la IA [3], para luego volver a perder tracción en los noventa. Cuál fue mi sorpresa cuando el año 2012 las redes neuronales artificiales pasaban a ser el método que todos querían usar y del cual todos hablaban. El motivo fue el sorprendente resultado del equipo SuperVision de la Universidad de Toronto[1] —Krizhevsky, Sutskever y Hinton—, que usando una red neuronal convolucional profunda (*deep convolutional neural network*) con 60 millones de parámetros y 650 mil neuronas, entrenado con dos GPUs durante una semana, ganaba el ImageNet challenge 2012 con un error top-5 del 15,3% y más de 10 puntos de mejora en relación al segundo lugar. Las redes neuronales profundas tenían algunos antecedentes importantes de buen rendimiento [4], pero el resultado del 2012 en el ImageNet challenge catapultó su popularidad. La arquitectura de red neuronal creada empezó a ser popularmente conocida como AlexNet [5], debido al nombre del primer autor, Alex Krizhevsky. A partir de ese momento, ingenieros e investigadores de diferentes áreas de la inteligencia artificial querían escribir los términos *deep learning*

1 | https://www.image-net.org/challenges/LSVRC/2012/results.html.

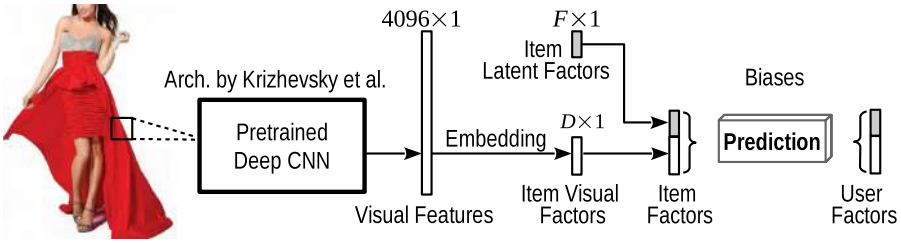


Fuente: [16].



Fuente: http://colah.github.io/posts/2015-08-Understanding-LSTMs/.

Figura 1. Arquitectura de GRU4Rec donde cada capa GRU tiene celdas GRU como la que se observa a la derecha, que pueden recordar y olvidar, selectivamente, permitiendo el aprendizaje de secuencias.



Fuente: [17].

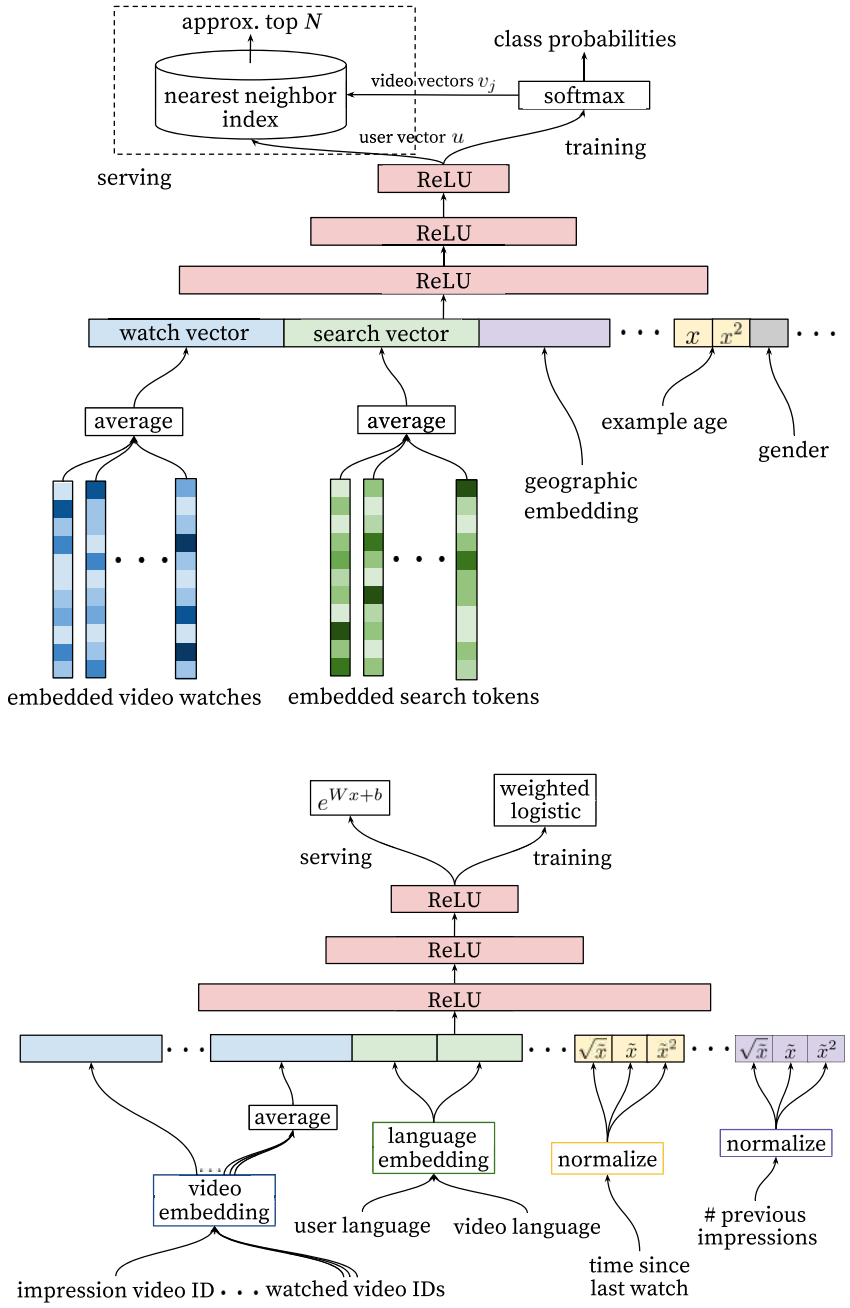
Figura 2. Diagrama de VBPR que indica cómo las características visuales obtenidas con una red neuronal convolucional profunda son incorporadas en el predictor de preferencia.

o *neural network* en el título de sus artículos, y es así cómo este método empieza a permear desde el campo de visión por computador a otras áreas como recuperación de información [6], traducción automática [7], describir imágenes con texto de forma automática [8], o incluso áreas creativas como generación visual [9] y musical [10].

A pesar del frenesí de distintas áreas por usar aprendizaje profundo, no fue hasta el 2015 que aparecen *papers* relevantes de aprendizaje profundo aplicados a Sistemas Recomendadores (de aquí en adelante, *SisRec*). Recordemos que los SisRec tienen como rol principal ayudarnos a encontrar ítems relevantes dentro de una sobreabundancia de información [11] considerando nuestras preferencias individuales. Compañías tan diversas como Amazon, Netflix, Google, Booking y Spotify basan buena parte de sus funcionalidades y modelos de negocio en sistemas recomendadores. Estos sistemas se han desarrollado por más de treinta años, pero han evolucionado especialmente rápido en la última década.

Volviendo a la aplicación de aprendizaje profundo aplicado a SisRec, es posible rescatar como antecedente previo a ImageNet el uso de *restricted Boltzman machines* [12], un tipo de red neuronal probabilística, entre los mejores métodos que compitieron en el Netflix prize [13]. Sin embargo, los primeros trabajos utilizando aprendizaje profundo ya sea a través de representaciones preentrenadas o para el modelo completo fueron los trabajos de Van den Oord *et al.* [14], un recomendador de música que utilizaba representaciones de audio aprendidas con una red neuronal profunda. Luego, se presenta en 2015 "aprendizaje profundo colaborativo para SisRec" [15], un método que combina las técnicas de filtrado colaborativo con *denoising autoencodeers*. El mismo 2015 aparece GRU4Rec [16] que modela secuencias de interacciones usando redes recurrentes con celdas GRU (ver Figura 1) para recomendar productos, y el mismo año se publica VBPR [17], método que utiliza la representación de imágenes que entrega una red convolucional preentrenada para mejorar recomendaciones visuales (ver Figura 2) realizadas por el modelo BPR [18].

Es difícil saber por qué el área de SisRec demoró tanto (alrededor de tres años) en ingresar a la ola de las redes neuronales profundas, pero es posible argumentar algunas razones en base a los pilares que posibilitaron el crecimiento del aprendizaje profundo: (a) gran cantidad de datos, (b) algoritmos de aprendizaje más eficientes, y (c) hardware especializado para el entrenamiento. En el área de sistemas de recomendación no era trivial encontrar *datasets* de gran tamaño, como el ImageNet, para entrenar modelos con tantos millones de parámetros como una red neuronal profunda. Esto se debe a que las grandes



Fuente: [26].

Figura 3. Las dos redes neuronales que formaban parte del sistema recomendador de videos, de aprendizaje profundo, del portal YouTube, activo hasta el 2019.

compañías han sido reticentes a compartir *datasets* que indiquen preferencias de usuarios por productos, ya sea por temas de competencia como para evitar violaciones de privacidad [19]. En los últimos años la disponibilidad de grandes *datasets* para entrenar modelos de recomendación ha mejorado mucho, con *datasets* como el de Spotify[2], Goodreads[3] o la versión 25M del tradicional movielens dataset[4]. En cuanto a algoritmos, si bien es posible adaptar métodos existentes de clasificación de imágenes o ranking de documentos para tareas de recomendación, el hecho de tener que incorporar el modelo de usuario en el método complejiza un poco su modelamiento e implementación. No es lo mismo usar un modelo de ranking de imágenes dada una imagen de entrada, que un modelo de ranking de imágenes personalizado, que considere tanto el historial de consumo de un usuario [17, 20, 21] así como el contexto de dicho consumo —día de la semana, hora, haciendo qué actividad, etc. [22]. En relación a hardware, no es un secreto que son grandes compañías como NVidia, Google, Amazon, o Facebook quienes disponen de los mejores recursos de hardware para entrenar modelos que crecen sin cesar en cantidad de parámetros: como muestra, el reciente modelo de lenguaje GPT-3 tiene 175 mil millones de parámetros [23], comparado con los 60 millones de parámetros de la AlexNet. Esto dificulta la investigación que provenga exclusivamente desde la academia, donde los incentivos permiten investigar temas diferentes a los que empujan la investigación en la industria. A pesar de estas dificultades, una propiedad interesante de estos modelos es la posibilidad de hacer *transfer learning* [24], es decir, entrenarlos inicialmente para una tarea y luego actualizar todos o parte de sus pesos para otro *dataset* o para otras tareas. Esto permite que el costo mayor de entrenamiento lo lleven a cabo grandes compañías, fundaciones y universidades,

2 | https://www.aicrowd.com/challenges/spotify-million-playlist-dataset-challenge.
3 | https://sites.google.com/eng.ucsd.edu/ucsdbookgraph/home.
4 | https://grouplens.org/datasets/movielens/25m/.

y luego otros usuarios con menores recursos de hardware tienen sólo que adaptar (*finetuning)* los pesos para la nueva tarea o *dataset* que se aborda.

A partir del año 2016 el aprendizaje profundo aterriza con fuerza en la conferencia internacional ACM de sistemas recomendadores, donde se publica "*Ask the GRU*" [25], un recomendador con aprendizaje multitarea de artículos científicos que usa una red recurrente con celdas del tipo Gated Recurrent Unit. Además de este *paper*, autores de Google [26] presentan la nueva versión del sistema recomendador de videos de YouTube, basado en dos redes neuronales profundas (ver Figura 3), una red que selecciona cientos de candidatos a partir de millones de opciones, y una segunda red que ordena los videos candidatos previamente filtrados. La nueva arquitectura del portal YouTube [27] tiene algunos aspectos interesantes, por ejemplo que considera los likes de los usuarios para generar el perfil del usuario para recomendar, cosa que no hacía el recomendador anterior [26].

Luego de estas publicaciones, es común encontrar SisRec implementados con métodos de aprendizaje profundo en temas como recomendación de música, películas, libros, pareja sentimental, ropa de temporada, entre muchos otros. Los sistemas han evolucionado en los últimos años de la arquitecturas como Transformer [28], integrados con otras técnicas como aprendizaje reforzado profundo [29], así como explotando avances en áreas como NLP [30] o modelos generativos [31].

Discusión y conclusión

El aprendizaje profundo ha impactado positivamente el área de SisRec, tanto como a otras áreas de aplicación de la inteligencia artificial. Hay, sin embargo, dos aspectos importantes a mencionar que generan inquietud en el área: cuánto es el progreso real que ha traído el aprendizaje profundo, y cómo estos modelos afectan el avance en temas de temas de equidad, explicabilidad y transparencia.[5]

¿Cuánto se ha progresado? El artículo de [32] pone en entredicho el impacto del aprendizaje profundo en los SisRec, mostrando que cuando métodos tradicionales de factorización matricial que se conocen por más de una década son entrenados adecuadamente, tienen tanto o mejor rendimiento que métodos de aprendizaje profundo. Si bien este *paper* es relevante por mostrar una crisis de reproducibilidad en SisRec y que no siempre el aprendizaje profundo puede mejorar el rendimiento los métodos ya conocidos, hay un aspecto relevante a considerar. La investigación de Dacrema sólo considera tuplas usuario-ítem como entrada, pero no considera información adicional como imágenes, video, metadata, contexto, etc. Justamente es con esta gran cantidad y diversidad de datos donde es esperable el rendimiento mejorado de técnicas de aprendizaje profundo, por lo cual se recomienda revisar con cautela los resultados de este análisis, y ponerlo en perspectiva sólo para el filtrado colaborativo tradicional.

FAccT. Considerar los desafíos que se plantean en la inteligencia artificial en relación a equidad (*fairness*), explicabilidad (*accountability*) y transparencia (*transparency*) es un gran desafío para los modelos de aprendizaje profundo en SisRec [33]. Considere el caso en que usa GPT-3, un modelo de 175 mil millones de parámetros, para recomendar un documento y el usuario solicita una explicación sobre dicha sugerencia ¿cómo explicaría dicha recomendación intentando ser transparente? Los métodos de explicabilidad para inteligencia artificial están en activa investigación en estos días [34] y si deseamos que los sistemas de recomendación permeen áreas críticas de toma de decisiones como medicina, finanzas o seguridad, se debe avanzar en esta área. En relación a asegurar que estos sistemas no están sesgados existe una inquietud similar: cómo hacer que provean recomendaciones justas a diferentes grupo de usuarios finales, por ejemplo de un sistema de recomendación de empleo, así como a creadores de contenido: que un portal de libros recomiende con la misma probabilidad tanto a escritores hombres como mujeres o de otros grupos LGBTQ.

Conclusión. El aprendizaje profundo tomó algunos años en permear el área de sistemas de recomendación en comparación con otras áreas de inteligencia artificial, pero se instaló con fuerza a partir de 2016 gracias a su gran capacidad para encontrar representaciones de usuarios y datos para posteriormente ser usadas en tareas de filtrado de información. Con el avance de modelos de visión por computador, modelos de lenguaje, arquitecturas como atención y más recientemente modelos de redes neuronales para grafos, el impacto de las redes neuronales profundas en SisRec no ha dejado de crecer. La integración de estas técnicas con otras como aprendizaje reforzado para SisRec y el crecimiento en los últimos años de los sistemas de recomendación conversacionales [35] le siguen dando fuerza a esta área de investigación. Los desafíos en términos de mostrar los avances reales en rendimiento de estas técnicas [32] así como su adaptación para lidiar con necesidades de equidad, transparencia, explicabilidad [33], nos harán ver sin duda mucha más investigación en este tema en los años venideros. ■

5 | FAccT 2018. ACM Conference on Fairness, Accountability, and Transparency https://facctconference.org/.

REFERENCIAS

[1] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *the Journal of machine Learning research*, *3*, 993-1022.

[2] Rosenblatt, F. (1957). *The perceptron, a perceiving and recognizing automaton Project Para*. Cornell Aeronautical Laboratory.

[3] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1985). *Learning internal representations by error propagation*. California Univ. San Diego La Jolla Inst. for Cognitive Science.

[4] Ciresan, D. C., Meier, U., Masci, J., Gambardella, L. M., & Schmidhuber, J. (2011). Flexible, high performance convolutional neural networks for image classification. In *Twenty-second international joint conference on artificial intelligence*.

[5] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*, 1097-1105.

[6] Severyn, A., & Moschitti, A. (2015). Learning to rank short text pairs with convolutional deep neural networks. In *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval* (pp. 373-382).

[7] Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.

[8] Vinyals, O., Toshev, A., Bengio, S., & Erhan, D. (2015). Show and tell: A neural image caption generator. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3156-3164).

[9] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial networks. arXiv preprint arXiv:1406.2661.

[10] Roberts, A., Engel, J., Raffel, C., Hawthorne, C., & Eck, D. (2018). A hierarchical latent vector model for learning long-term structure in music. In *International Conference on Machine Learning* (pp. 4364-4373). PMLR.

[11] McNee, S. M., Kapoor, N., & Konstan, J. A. (2006). Don't look stupid: avoiding pitfalls when recommending research papers. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work* (pp. 171-180). ACM.

[12] Salakhutdinov, R., Mnih, A., & Hinton, G. (2007). Restricted Boltzmann machines for collaborative filtering. In Proceedings of the 24th international conference on Machine learning (pp. 791-798).

[13] Bennett, J., & Lanning, S. (2007, August). The Netflix Prize. In Proceedings of KDD cup and workshop (Vol. 2007, p. 35).

[14] Van Den Oord, A., Dieleman, S., & Schrauwen, B. (2013). Deep content-based music recommendation. In Neural Information Processing Systems Conference (NIPS 2013) (Vol. 26). Neural Information Processing Systems Foundation (NIPS).

[15] Wang, H., Wang, N., & Yeung, D. Y. (2015). Collaborative deep learning for recommender systems. In Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1235-1244).

[16] Hidasi, B., Karatzoglou, A., Baltrunas, L., & Tikk, D. (2015). Session-based recommendations with recurrent neural networks. arXiv preprint arXiv:1511.06939.

[17] He, R., & McAuley, J. (2016). VBPR: visual bayesian personalized ranking from implicit feedback. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 30, No. 1).

[18] Rendle, S., Freudenthaler, C., Gantner, Z., & Schmidt-Thieme, L. (2012). BPR: Bayesian personalized ranking from implicit feedback. arXiv preprint arXiv:1205.2618.

[19] Narayanan, A., & Shmatikov, V. (2006). How to break anonymity of the Netflix Prize dataset. arXiv preprint cs/0610105.

[20] Chen, J., Zhang, H., He, X., Nie, L., Liu, W., & Chua, T. S. (2017). Attentive collaborative filtering: Multimedia recommendation with item-and component-level attention. In Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval (pp. 335-344).

[21] Messina, P., Domínguez, V., Parra, D., Trattner, C., & Soto, A. (2019). Content-based artwork recommendation: integrating painting metadata with neural and manually-engineered visual features. User Modeling and User-Adapted Interaction, 29(2), 251-290.

[22] Adomavicius, G., & Tuzhilin, A. (2011). Context-aware recommender systems. In Recommender systems handbook (pp. 217-253). Springer, Boston, MA.

[23] Brown, T. B., Mann, B., Ryder, N., et al. (2020). Language models are few-shot learners. arXiv preprint arXiv:2005.14165

[24] Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. IEEE Transactions on knowledge and data engineering, 22(10), 1345-1359.

[25] Bansal, T., Belanger, D., & McCallum, A. (2016). Ask the gru: Multi-task learning for deep text recommendations. In proceedings of the 10th ACM Conference on Recommender Systems (pp. 107-114).

[26] Covington, P., Adams, J., & Sargin, E. (2016). Deep neural networks for YouTube recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems* (pp. 191-198). ACM.

[27] Zhao, Z., Hong, L., Wei, L. et al. (2019). Recommending what video to watch next: a multitask ranking system. In Proceedings of the 13th ACM Conference on Recommender Systems (pp. 43-51).

[28] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. arXiv preprint arXiv:1706.03762.

[29] Zheng, G., Zhang, F., Zheng, Z., Xiang, Y., Yuan, N. J., Xie, X., & Li, Z. (2018, April). DRN: A deep reinforcement learning framework for news recommendation. In Proceedings of the 2018 World Wide Web Conference (pp. 167-176).

[30] Penha, G., & Hauff, C. (2020). What does BERT know about books, movies and music? Probing BERT for Conversational Recommendation. In Fourteenth ACM Conference on Recommender Systems (pp. 388-397).

[31] Kang, W. C., Fang, C., Wang, Z., & McAuley, J. (2017). Visually-aware fashion recommendation and design with generative image models. In 2017 IEEE International Conference on Data Mining (ICDM) (pp. 207-216). IEEE.

[32] Dacrema, M. F., Cremonesi, P., & Jannach, D. (2019). Are we really making much progress? A worrying analysis of recent neural recommendation approaches. In *Proceedings of the 13th ACM Conference on Recommender Systems* (pp. 101-109).

[33] Ekstrand, M. D., & Sharma, A. (2017). FATREC Workshop on Responsible Recommendation. In Proceedings of the Eleventh ACM Conference on Recommender Systems (pp. 382-383).

[34] Gunning, D. (2017). Explainable artificial intelligence (xai). Defense Advanced Research Projects Agency (DARPA), nd Web, 2(2).

[35] Christakopoulou, K., Radlinski, F., & Hofmann, K. (2016). Towards conversational recommender systems. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 815-824).

Detección de discurso de odio

AYMÉ ARANGO Estudiante de Doctorado del Departamento de Ciencias de la Computación de la Universidad de Chile

Las redes sociales se han convertido en un medio importante de interacción entre usuarios de todo el mundo. El contenido compartido puede ser de gran utilidad, como fuente de información inmediata que permite el análisis de eventos, estudio de fenómenos, la difusión de arte, ciencia, entre otras. Junto con esta información, también se encuentran manifestaciones de ciertos fenómenos comunicacionales como noticias falsas y discurso de odio que pueden producir efectos colaterales dañinos.

A pesar de que hay cierta discrepancia en cómo definir el término "discurso de odio", una de las definiciones más usadas es: expresiones derogatorias a individuos o grupos atendiendo a cierta característica como color de la piel, origen étnico, género, orientación sexual, entre otros.[1] La propagación de este tipo de contenido en los medios digitales tiene como efectos la molestia e intimidación de los usuarios. En casos extremos puede trascender el ámbito virtual y llegar a ocasionar daños físicos en individuos. Estudios recientes han encontrado vínculos entre el odio en las redes y los crímenes de odio [1]. Desde diversas disciplinas se trabaja para entender y tratar de identificar a tiempo este fenómeno.

Revisar el contenido publicado consiste en una ardua tarea para los proveedores de redes sociales. Debido al gran flujo de datos a analizar en un red social, y a su variedad, se requieren técnicas automatizadas para detectar este tipo de contenido y tomar medidas necesarias a tiempo. Dada la complejidad de la tarea, esto no ha podido lograrse satisfactoriamente hasta el momento.

Desde el punto de vista de la ciencia de datos, la detección de discurso de odio puede ser planteada como un problema de clasificación en el cual la entrada es un mensaje (*tweet*, comentario, fotografía, etc.) y la salida es la clasificación de éste como contenido odioso o no. Sin embargo, algunos investigadores consideran categorías más específicas y construyen modelos capaces de predecir el tipo específico de odio que está siendo expresado, como sexismo, racismo, xenofobia, entre otros.

Técnicas de inteligencia artificial se han venido utilizando para intentar resolver este problema. Específicamente, los modelos de aprendizaje automático han sido ampliamente utilizados como herramientas en la detección de discurso de odio [2, 3], incluyendo, en los últimos años, modelos basados en arquitecturas de redes neuronales [4]. Para que tales modelos "aprendan" a diferenciar el contenido "odioso" del contenido "normal", se necesitan datos previamente etiquetados. Idealmente, estos datos deberían contener ejemplos representativos de los diferentes tipos de expresiones de odio existentes. Obtener este tipo de datos etiquetados es costoso y debido a la información sensible que manejan y a políticas de cada

1 | https://www.encyclopedia.com/international/encyclopedias-almanacs-transcripts-and-maps/hate-speech.

plataforma, muy pocos conjuntos de datos son públicos y la mayoría son pequeños.[2] Adicionalmente, algunos de los conjuntos de datos publicados han sido reportados como sesgados [5], lo que reduce las posibilidades de utilizar datos de calidad, y como consecuencia, de construir buenos detectores de discurso de odio.

Como parte de mi tesis doctoral, junto con los profesores Bárbara Poblete y Jorge Pérez, estamos investigando técnicas para la construcción de modelos que sean generalizables a diferentes idiomas. Tal y como sucede en otras tareas relacionadas con el Procesamiento del Lenguaje Natural, la mayoría de los modelos desarrollados hasta el momento han sido principalmente explotados para resolver el problema en el idioma inglés. Como consecuencia, la gran parte de los recursos construidos son de utilidad solamente para este idioma, mientras la tarea avanza más lentamente para el resto. Analizando dos de los mejores modelos reportados en la literatura de idioma Inglés [6], encontramos que los resultados mostrados estaban sobreestimados debido a problemas experimentales, y uso de datos sesgados. Además, estos modelos presentan una pobre generalización a datos en el mismo idioma inglés y a datos en español.

Siendo el odio en medios digitales un fenómeno del cual hay evidencia a lo largo de todo el mundo, se requieren soluciones efectivas en los distintos idiomas para afrontar el problema. La idea de nuestro enfoque es aprovechar los recursos existentes (mayormente en inglés) y construir modelos generalizables a diferentes idiomas, ahorrando así el esfuerzo necesario en la creación de nuevos recursos para cada idioma separadamente. Para que los modelos de aprendizaje automático sean capaces de transferir conocimiento de un idioma a otro, se requieren representaciones de los datos a través de un conjunto de características que puedan ser comunes para diferentes idiomas. Ejemplo de esto pueden ser representaciones vectoriales multilingües o información que no esté directamente relacionada con un idioma específico. Particularmente, nuestro equipo de investigación ha trabajado en encontrar dichas características que sean comunes al odio en diferentes idiomas que nos permitan construir modelos generalizables. Bajo nuestro foco de atención, se encuentran aquellas representaciones que puedan ser extraídas del contexto del mensaje, del autor del mensaje (meta-información) y que por su naturaleza no estén atadas a un único idioma [7]. Además, estamos interesados en construir representaciones específicas para el lenguaje de odio, siendo este un fenómeno con características especiales donde ciertas palabras o expresiones pueden tomar connotaciones de odio, en dependencia del contexto. Dichas expresiones no son únicas y pueden depender no sólo del idioma, sino del contexto cultural en el que se exprese. Nos interesaría resaltar estas diferencias culturales en aras de construir modelos que generalicen mejor.

Este tipo de generalización presenta aún varios retos debido a las diferentes características de los idiomas y a la complejidad que puede tener la tarea, siendo el odio un fenómeno no sólo lingüístico, sino social y cultural. Definitivamente, todavía hay mucho que investigar en esta área. Los resultados aún no son concluyentes respecto a qué modelo o representación de datos resulta mejor para esta tarea y aunque se han logrado algunos avances, la tarea aún está por resolverse. ■

REFERENCIAS

[1] Williams ML, Burnap P, Javed A, Liu H, Ozalp S. Hate in the machine: anti-black and anti-Muslim social media posts as predictors of offline racially and religiously aggravated crime. Br J Criminol (2020), 60(1), pp. 93–117.

[2] Anzovino, M., Fersini, E., and Rosso, P. Automatic Identification and Classification of Misogynistic Language on Twitter. In International Conference on Applications of Natural Language to Information Systems (2018), Springer, pp. 57–64.

[3] Papegnies, E., Labatut, V., Dufour, R., and Linares, G. Graph-based Features for Automatic Online Abuse Detection. In International Conference on Statistical Language and Speech Processing (2017), Springer, pp. 70–81.

[4] Gambäck, B., and Sikdar, U. K. Using Convolutional Neural Networks to Classify Hate-Speech. In Proceedings of the First Workshop on Abusive Language Online (2017), Association for Computational Linguistics, pp. 85–90.

[5] Maarten Sap, Dallas Card, Saadia Gabriel, Yejin Choi, and Noah A. Smith. The Risk of Racial Bias in Hate Speech Detection. In Proceedings of the Association for Computational Linguistics (2019), pp. 1668–1678.

[6] Arango, A., Pérez, J., Poblete, B.: Hate Speech Detection is Not as Easy as You May Think: A Closer Look at Model Validation. In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (2019), ACM, pp. 45–54.

[7] Arango, A., Pérez, J., & Poblete, B. Hate Speech Detection is Not as Easy as You May Think: A Closer Look at Model Validation (extended version). Information Systems , 101584 (2020).

2 | https://github.com/aymeam/Datasets-for-Hate-Speech-Detection.

Conectando la visión y el lenguaje

JESÚS PÉREZ-MARTÍN Estudiante de Doctorado del Departamento de Ciencias de la Computación de la Universidad de Chile e Investigador del Instituto Milenio Fundamentos de los Datos.

BENJAMÍN BUSTOS Profesor Titular del Departamento de Ciencias de la Computación de la Universidad de Chile e Investigador Asociado del Instituto Milenio Fundamentos de los Datos.

JORGE PÉREZ Profesor Asociado del Departamento de Ciencias de la Computación de la Universidad de Chile e Investigador Asociado del Instituto Milenio Fundamentos de los Datos.

En este minuto más de 500 horas de video se están publicando en YouTube.[1] Además, el último *Digital Global Overview Report* estima que diariamente se visualizan mil millones de horas de video en la misma plataforma. Con los videos ganando tanta popularidad, YouTube Creator Academy[2] recomienda que las descripciones transmitan información valiosa para ayudar a los espectadores a encontrar videos en los resultados de búsquedas y comprender lo que mirarán.[3] En este sentido detalla: *"Las descripciones bien redactadas con las palabras clave correctas pueden ayudar a mejorar las visualizaciones y el tiempo de reproducción, ya que ayudan a que el video tenga una mayor visibilidad en los resultados de la búsqueda"*.

La forma de comunicación que más usamos los humanos es el lenguaje natural. Es entonces esencial que sistemas interactivos de Inteligencia Artificial (IA) y robots auxiliares sean capaces de generar texto automáticamente a partir de datos no lingüísticos. Reiter y Dale [1] caracterizan *Natural Language Generation* (NLG) como la producción de textos comprensibles a partir de una representación no lingüística subyacente de la información. Esta definición de NLG generalmente se asocia con la de *data-to-text generation*, asumiendo que la entrada exacta puede variar sustancialmente.

Hoy en día, la generación de texto a partir de una entrada perceptiva no estructurada —como una imagen sin

1 | Estadísticas de YouTube 2021 [infografía] - 10 datos fascinantes de YouTube: https://cl.oberlo.com/blog/estadisticas-youtube.

2 | Academia de creadores de YouTube, educación y cursos: https://creatoracademy.youtube.com.

3 | Consejos de YouTube para crear descripciones inteligentes: https://creatoracademy.youtube.com/page/lesson/descriptions?hl=es-419#strategies-zippy-link-1.

procesar o un video— se ha convertido en un desafío importante en el campo de investigación reciente que combina Visión y Lenguaje (V+L). Específicamente, obtener texto a partir de un video (*video-to-text*) puede efectuarse, principalmente, recuperando las descripciones más significativas de un corpus o generando una nueva descripción dado el video de contexto. Estas dos formas representan tareas esenciales para las comunidades de procesamiento de lenguaje natural y visión computacional, y son ampliamente conocidas como *video-to-text retrieval* y *video captioning/description*, respectivamente. Ambas tareas son sustancialmente más complejas que generar o recuperar una oración desde una única imagen. La información espacio-temporal presente en los videos introduce diversidad y complejidad respecto al contenido visual y a la estructura de las descripciones de lenguaje asociadas.

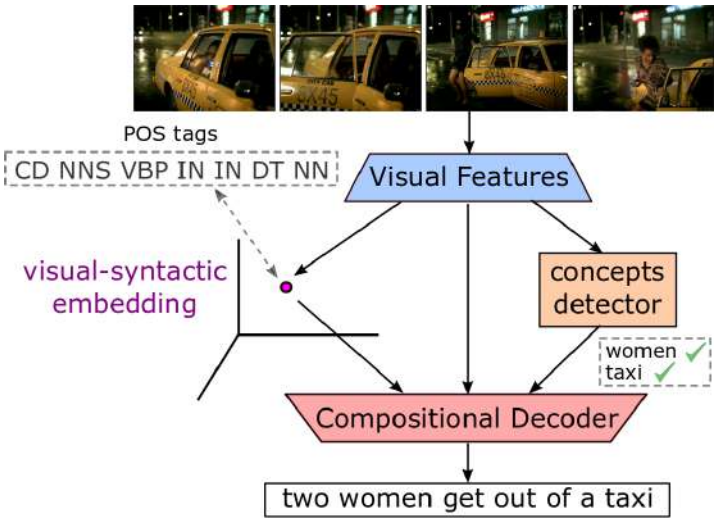


Figura 1. *Video captioning* usando un *embedding visual-sintáctico*. El método obtiene representaciones semánticas y sintácticas de alto nivel a partir de la representación visual del video. A continuación, el decodificador genera una oración a partir de ellos.

Con gran atención de ambas comunidades, V+L incluye otras tareas desafiantes que conectan o combinan las modalidades de la visión y el lenguaje, como *visual question-answering* (responder preguntas basadas en texto sobre imágenes), *caption-based image/video retrieval* (dados un texto y un grupo de imágenes, debemos recuperar la imagen que mejor se describe con el texto), *video generation from text* (generar un video plausible y diverso a partir de un texto de entrada) y *multimodal verification* (dada una o más imágenes y un texto, debemos predecir alguna relación semántica).

Sintaxis y semántica de un video

Es impresionante el progreso que los investigadores han logrado en conjuntos de datos específicos, pero a pesar de este progreso, la conversión de video a texto sigue siendo un problema abierto. Las técnicas del estado del arte aún están lejos de lograr un desempeño similar al humano. No obstante, las técnicas basadas en *deep learning* han logrado resultados prometedores, tanto para la generación de descripciones como para los métodos basados en la recuperación.

Como una tarea de generación de texto, el proceso de describir videos requiere predecir una secuencia de palabras semántica y sintácticamente correcta dado el contexto presente en el video. Los primeros trabajos en esta área siguieron la estrategia de, primero, detectar sujeto, verbo y objeto, formando un *triplete SVO*; y luego, generar una oración usando un conjunto reducido de plantillas que aseguran la correctitud gramatical. Este enfoque requiere que los modelos reconozcan a los sujetos y objetos que participan en la acción que debemos describir, logrando sus mejores resultados en videos cortos de entornos específicos, como deporte o cocina. En este tipo de videos, la cantidad de objetos y acciones que se debe detectar es limitada.

A partir de esta idea, podemos notar que para los modelos de *video captioning* dos aspectos esenciales son la identificación de contenidos visuales de forma explícita y la intención de producir oraciones correctas. Desarrollar técnicas que aborden alguno de estos aspectos ha guiado la investigación en los últimos años. Por un lado tenemos métodos que intentan conectar las palabras generadas a regiones específicas dentro del video (*visual grounding*) [2] y modelar las relaciones entre ellas [3, 4]. Mientras que por el otro tenemos métodos que consideran el aprendizaje de una representación sintáctica como un componente esencial de los enfoques de *video captioning* [5, 6, 7].

En el Departamento de Ciencias de la Computación (DCC) de la Universidad de Chile nos encontramos desarrollando métodos de *video captioning* que extraen información valiosa sobre las posibles descripciones a partir de dimensiones implícitas en la información visual. Nuestros resultados recientes muestran que los videos contienen, además de la apariencia y el movimiento, información semántica y sintáctica que podemos extraer directamente de la información visual para guiar el proceso de generación de

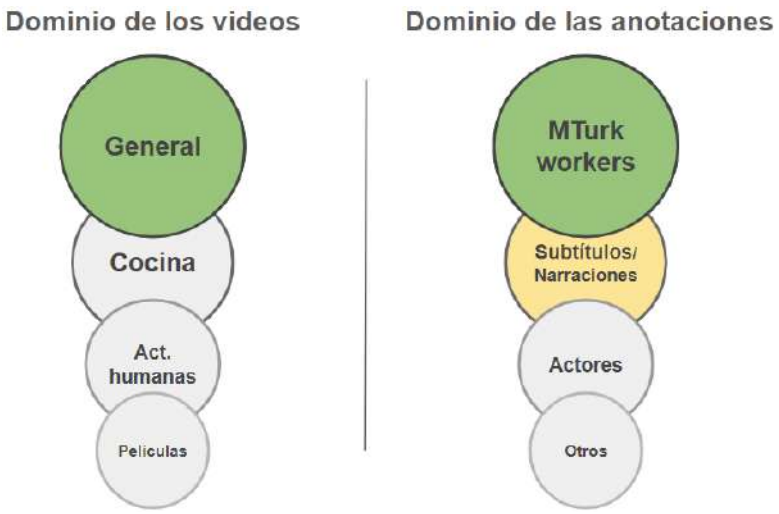


Figura 2. Para entrenar estos métodos, existen más de veinticinco conjuntos de datos anotados que podemos agrupar según el dominio de video y de diferentes formas se obtienen las descripciones.

texto. Sin embargo, tener una fuerte dependencia de sólo una de ellas puede perjudicar el rendimiento de los modelos, produciendo brechas semánticas u oraciones sintácticamente incorrectas. Por eso, para nosotros es fundamental determinar cómo fusionar estos canales de información de forma adaptativa. En dos artículos que presentamos recientemente en las conferencias internacionales ICPR 2020 [8] y WACV 2021 [7], proponemos estrategias efectivas que combinan técnicas de recuperación y generación para evitar estas brechas y aprender representaciones de forma multimodal.

Específicamente, en nuestro trabajo propusimos un modelo llamado *Visual-Semantic-Syntactic Aligned Network* (SemSynAN) [7]. Este modelo basado en el esquema *encoder-decoder* es capaz de generar oraciones con semántica y sintaxis más precisas. Una de las innovaciones más importante fue proponer una técnica de recuperación de secuencias de etiquetado gramatical (POS por sus siglas en inglés)[4] provenientes de las descripciones de video, para generar representaciones sintáctica de alto nivel directamente desde la información visual (ver Figura 1). Con este trabajo mostramos que prestar atención especial a la sintaxis puede mejorar sustancialmente la calidad de las descripciones. Además, nuestro método garantiza la relación contextual entre las palabras de la oración, controlando el significado semántico y la estructura sintáctica de las descripciones generadas [7].

Conjuntos de datos de entrenamiento

V+L es un área de investigación recientemente planteada. Aunque ha recibido mucha atención en los últimos años, todavía se necesitan más datos para entrenar y evaluar nuevos modelos. Para distinguir con precisión entre diferentes clases de información visual, los modelos deben entrenarse a escala, con descripciones diversas y de alta calidad que contengan una amplia variedad de videos.

La creación de conjuntos de datos a gran escala requiere un esfuerzo humano significativo y costoso para su anotación, ya que recopilar una gran cantidad de referencias puede llevar mucho tiempo y ser difícil para los idiomas menos comunes. Debido a esto —y a pesar de que la mayor cantidad de *datasets* ha sido creada a partir de videos de dominio general anotados por humanos (ver Figura 2)—, el *dataset* más grande a la fecha ha sido creado a partir de la generación automática de subtítulos y narraciones (*dataset HowTo100M* [9]).

Con trabajos recientes como CLIP [10], el campo se ha movido a nuevas arquitecturas y modelos (*transformers* [11], *pre-training* y *fine-tuning* ahora se han convertido en el enfoque dominante). Básicamente, estos estudios han mostrado los beneficios de preentrenar los modelos para tareas de V+L y luego ajustar el modelo para tareas específicas.

Por ejemplo, podemos aprender previamente representaciones genéricas a partir de tareas de V+L, como *visual question-answering* o *cross-modal retrieval* (recuperación a través de diferentes modalidades, como imagen-texto, video-texto y audio-texto), y luego ajustar su codificación visual en la tarea de *video captioning*. Esta técnica requiere un gran volumen de datos para aprender dicha representación en un espacio común entre la información visual y textual. Por ejemplo, para entrenar CLIP se usaron 400 millones de pares (imagen, texto) obtenidos de Internet.

Los modelos de *video captioning* basados en esta estrategia, como COOT [12],

4 | Categorizar y etiquetar palabras de acuerdo a categorías léxicas: https://www.nltk.org/book/ch05.html.

generalmente son preentrenados sobre datos obtenidos de forma automática de los subtítulos y narraciones (ver Figura 2) que brindan las plataformas de video *online*. Sin embargo, un gran inconveniente de este tipo de corpus es la gran cantidad de *tokens* desconocidos (términos que no se pueden asociar a una palabra del vocabulario) que se producen. Por ejemplo, en *HowTo100M* [9] sólo el 36,64% de las palabras del vocabulario (217.361 de las 593.238 palabras únicas) aparecen en el vocabulario ampliamente utilizado *GloVe-6B*[5] [13], que tiene 400.000 *tokens*. Este alto nivel de "ruido" en los subtítulos es un aspecto interesante del proceso de entrenamiento que debemos aprender a aprovechar.

Conclusiones

Hace diez años pocos hubieran imaginado que sistemas de V+L serían capaces de generar descripciones textuales plausibles como las que se logran hoy. Los investigadores han logrado modelos que extraen, hasta cierto sentido, información espacio-temporal compleja presente en los videos. No obstante, una característica de la que carecen los sistemas actuales es la capacidad de representar el *sentido común*, por lo que aún queda mucho para comprender y representar la diversidad en cuanto a contenido visual de los videos y la estructura de sus descripciones textuales.

Es muy probable que en el futuro la cantidad de videos que los buscadores deberán procesar sea mayor que en la actualidad. Siempre ha sido así y al día de hoy, que la pandemia nos incita a ser más digitales, no hay ningún indicador que señale que esta dinámica cambiará. Al contrario, esta tendencia aumentará la necesidad de transformar la información visual en descripciones textuales que la resuman, verbalicen y simplifiquen de forma precisa. ▪

REFERENCIAS

[1] Reiter, E. & Dale, R. Building natural language generation systems. (Cambridge University Press, 2000).

[2] Pan, B. et al. Spatio-Temporal Graph for Video Captioning with Knowledge Distillation. In Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 10870–10879 (2020).

[3] Zhou, L., Kalantidis, Y., Chen, X., Corso, J. J. & Rohrbach, M. Grounded Video Description. In Proc. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 6571–6580 (IEEE, 2019).

[4] Zhang, Z. et al. Object Relational Graph with Teacher-Recommended Learning for Video Captioning. In Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 13278–13288 (2020).

[5] Hou, J., Wu, X., Zhao, W., Luo, J. & Jia, Y. Joint Syntax Representation Learning and Visual Cue Translation for Video Captioning. In Proc. IEEE International Conference on Computer Vision (ICCV) (2019).

[6] Wang, B. et al. Controllable Video Captioning with POS Sequence Guidance Based on Gated Fusion Network. In Proc. IEEE International Conference on Computer Vision (ICCV) (2019).

[7] Pérez-Martín, J., Bustos, B. & Pérez, J. Improving Video Captioning with Temporal Composition of a Visual-Syntactic Embedding. In Proc. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) (2021).

[8] Pérez-Martín, J., Bustos, B. & Pérez, J. Attentive Visual Semantic Specialized Network for Video Captioning. In Proc. 25th International Conference on Pattern Recognition (2020).

[9] Miech, A. et al. HowTo100M: Learning a Text-Video Embedding by Watching Hundred Million Narrated Video Clips. In Proc. IEEE/CVF International Conference on Computer Vision (ICCV) 2630–2640 (IEEE, 2019).

[10] Radford, A. et al. Learning Transferable Visual Models From Natural Language Supervision. (2021).

[11] Vaswani, A. et al. Attention is all you need. In Proc. 31st International Conference on Neural Information Processing Systems 6000–6010 (Curran Associates Inc., 2017).

[12] Ging, S., Zolfaghari, M., Pirsiavash, H. & Brox, T. COOT: Cooperative Hierarchical Transformer for Video-Text Representation Learning. In Proc. Conference on Neural Information Processing Systems (2020).

[13] Pennington, J., Socher, R. & Manning, C. D. Glove: Global vectors for word representation. IN EMNLP (2014).

5 | Proyecto Stanford GloVe (vectores globales) que usa aprendizaje no supervisado para obtener vectores representativos para un gran conjunto de palabras: https://nlp.stanford.edu/projects/glove/.

¿Cómo la inteligencia artificial puede ayudar al e-commerce?

EQUIPO IMPRESEE

CAMILA ÁLVAREZ Chief Technology Officer (CTO)
JUAN MANUEL BARRIOS Chief Executive Officer (CEO)
MAURICIO PALMA LIZANA Chief Financial Officer (CFO)
JOSÉ M. SAAVEDRA Chief Research Officer (CRO)

El *e-commerce* es un mercado mundial que se ha vuelto indispensable en el último tiempo. Basa su éxito en la satisfacción de los usuarios que necesitan comprar y en el consecuente incremento de las ventas en las tiendas. Es un contexto en el que modelos de Inteligencia Artificial (IA) y Ciencia de Datos se vuelven cada vez más relevantes tanto para atraer visitantes, mostrar productos relevantes, diseñar campañas de marketing, etc.

Impresee es una empresa SaaS que ofrece servicios de alta tecnología para el *e-commerce*. Tenemos clientes en diversas partes del mundo como Estados Unidos, Canadá, Alemania, China y Sudamérica, entre otras. Fundamos Impresee con el deseo de desarrollar servicios que combinen áreas de inteligencia artificial, visión por computadora, procesamiento del lenguaje natural y ciencia de datos para lograr soluciones innovadoras que mejoren el *e-commerce*.

La investigación científica la hacemos en Impresee eCommerce Labs[1], donde trabajamos en conjunto con retailers y colaboradores académicos para hacer investigación aplicada para el *e-commerce* y crear tecnología novedosa usando datos reales de ambientes reales. Nos enorgullece haber sido reconocidos por la comunidad científica en el año 2015 con el Premio a Mejor Demo basada en Visión por Computadora en la IEEE International Conference on Computer Vision (ICCV).

IA en la industria del *e-commerce*

En un principio nos enfocamos principalmente en mejorar la experiencia de los consumidores a través de un motor de búsqueda moderno, eficiente y efectivo. Potenciamos la tradicional búsqueda por texto con modelos basados en visión por computadora para permitir la búsqueda de productos por medio de fotos. Además desarrollamos una novedosa modalidad de consulta: la búsqueda basada en dibujos (*sketch-based image retrieval*), que tiene sus raíces en la tesis de doctorado

1 | https://impresee.com/ecommerce-labs/.

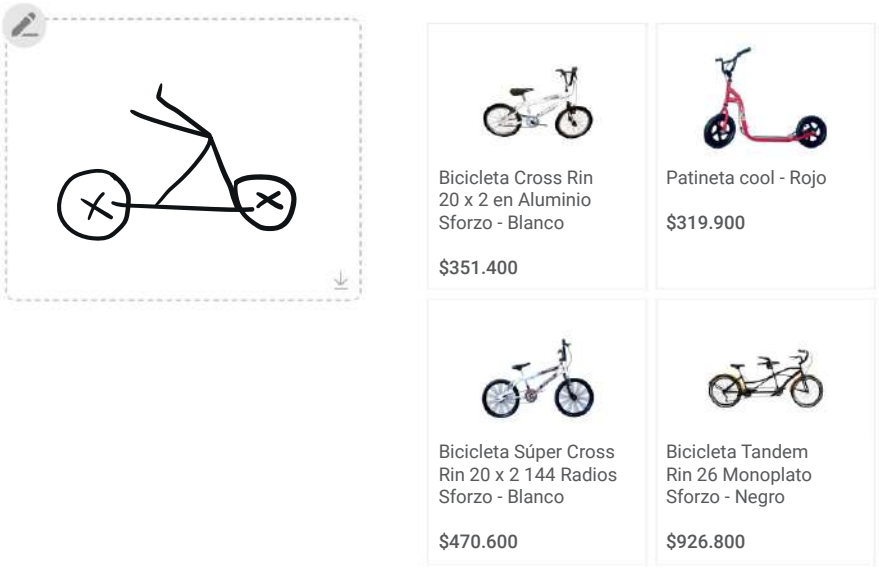


Figura 1. Resultado de búsqueda a través de dibujos.

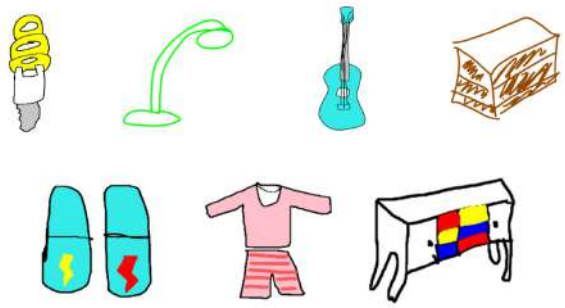


Figura 2. Ejemplo de consultas tipo sketch con color.

de José M. Saavedra. Por ejemplo, la Figura 1 muestra el resultado de búsqueda de un dibujo.

Luego observamos que la gran cantidad de datos que capturamos de una tienda (tráfico, visitantes, ventas) y los datos que generamos desde las búsquedas (consultas, fotos, dibujos, clicks) se complementan para formar un conjunto valioso para distintas áreas de la tienda. Trabajamos en crear métodos para analizar datos y generar información útil para la tienda, como el comportamiento de los visitantes y su apreciación de los productos para apoyar las áreas de marketing y ventas.

Nos dimos cuenta que los *dashboards* no son suficientes para generar valor, sino que debemos ir más allá, apoyando las conclusiones y automatizando las acciones posteriores. Por ejemplo, mediante análisis de datos es posible localizar productos con un buen potencial de ventas y que tienen baja visibilidad. Luego con *machine learning* es posible generar modelos para identificar las mejores acciones de marketing a realizar en una tienda para aumentar sus ventas.

Los sistemas de recomendación son otra arista que estamos trabajando. En esta línea investigamos modelos para integrar recomendadores y buscadores. En el buscador el visitante escribe lo que desea comprar y, además, en nuestro caso, puede subir una foto o dibujarlo. Según nuestros análisis, es tres veces más probable que un usuario que usa el buscador compre un producto comparado con uno que solo navega por el sitio. Por tanto, analizando la gran cantidad de imágenes de un catálogo (fotos de *influencers*, catálogos de temporada, etc.) junto con las imágenes de búsqueda, es posible entrenar modelos basados en redes convolucionales que permitan recomendar de forma automática prendas de vestir, dada una prenda de consulta. En términos técnicos, se trata de modelar un espacio de características donde las prendas complementarias se acercan entre sí.

Trabajos de investigación recientes

Trabajar en investigación en casos reales nos permite detectar problemas anticipadamente y desarrollar soluciones que tienen alto impacto. Así, en los siguiente párrafos describiremos tres trabajos aceptados para presentación oral en workshops de la International Conference on Computer Vision and Pattern Recognition (CVPR) 2021.

Color-Sketch-based Image Retrieval

Luego de lanzar el buscador basado en dibujos, observamos que en contextos como Fashion & Apparel y Home-Decor los usuarios debieran poder agregar información a la consulta como color y texturas. Así comenzamos investigar sobre cómo modelar dibujos incluyendo color y texturas y cómo compararlos con las imágenes de productos. La Figura 2 muestra algunas consultas. El resultado se plasmó en el

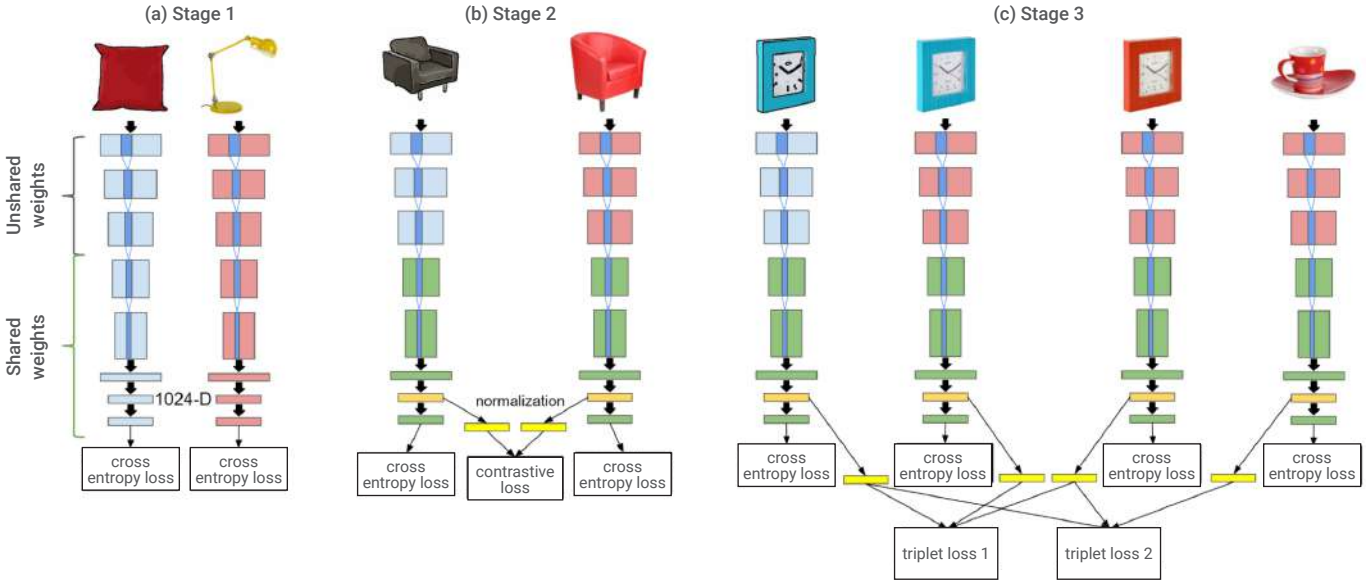


Figura 3. Arquitectura Sketch-QNet.

Figura 4. Ejemplo de resultado en un espacio de características de 8 dimensiones.

trabajo titulado "Sketch-QNet: A Quadruplet ConvNet for Color Sketch-based Image Retrieval"[2], que fue aceptado recientemente en el 1st Workshop on Sketch-Oriented Deep Learning (SketchDL) de CVPR 2021.

En ese trabajo proponemos una nueva arquitectura de red neuronal convolucional a la que llamamos Sketch-QNet para resolver el problema de *color-sketch based image retrieval*. La Figura 3, muestra la arquitectura propuesta que es entrenada por medio de cuadrupletas (cuatro pares de entrada). Con esto, extendemos la búsqueda de imágenes basada en dibujos a consultas que incluyan información de color. El objetivo es generar un espacio de características que pueda contener sketches con color y fotografías al mismo tiempo. El entrenamiento se realiza de modo que una consulta en forma de *sketch* con color quede muy cerca, en el espacio inducido, de fotos que expresen la misma información semántica de la consulta. Fotos que comparten solamente el concepto pero difieren en color deben quedar un poco más lejos. Finalmente, fotos con una semántica diferente a la consulta deben estar mucho más lejos de ella.

Representaciones compactas para Sketch-based Image Retrieval

La eficiencia de los espacios de características juegan un rol muy importante en sistemas reales. Comúnmente los vectores característicos para la recuperación de imágenes son de alta dimensión, variando entre 256 a 4096 dimensiones. Esto resulta impráctico para soportar catálogos con millones de imágenes, impactando negativamente el tiempo de búsqueda y la memoria requerida. Decidimos investigar modelos que nos permitan crear espacios reducidos (por ejemplo, menos de 10 dimensiones) sin perder efectividad. En esta línea desarrollamos el trabajo titulado "Compact and Effective Representations for Sketch-based Image Retrieval"[3], recientemente aceptado en el 1st Workshop on Sketch-Oriented Deep Learning (SketchDL) de CVPR 2021.

2 | https://impresee.com/sketch-qnet/.

3 | https://impresee.com/sketch-based-image-retrieval/.

En este trabajo, observamos que los espacios de características actuales forman una topología local que puede ser aprovechada por métodos de reducción de dimensión que preserven la localidad. Nuestros experimentos muestran que el uso de UMAP como método de reducción permite obtener espacios de baja dimensión (por ejemplo, 4 u 8) incrementando, además, la efectividad del método original. Este incremento en la efectividad se debe a que al preservar la localidad se extraen características relevantes a la vecindad de cada punto, descartando características ruidosas. Así, objetos que comparten una semántica similar tienden a ser atraídos entre sí. La Figura 4 muestra algunos resultados de recuperación de imágenes usando *sketches*, en un espacio reducido a 8 dimensiones. Estos resultados representan un nuevo estado del arte en este contexto.

Extracción de atributos visuales

Los atributos visuales juegan un rol muy importante en la búsqueda de productos. La manera tradicional de extraer estos atributos es entrenando una red CNN que se ajusta a un conjunto determinado de clases. Esta aproximación no escala a problemas donde los atributos de interés pueden cambiar con frecuencia. En nuestro trabajo titulado "Scalable Visual Attribute Extraction through Hidden Layers of a Residual ConvNet"[4] proponemos un método para extraer atributos visuales de imágenes, particularmente como las que podemos encontrar en un *e-commerce*, aprovechando la capacidad que tienen las capas ocultas de una red convolucional para aprender características visuales (ver Figura 5).

(a) Color (b) Textura (b) Color y Textura

Figura 5. Agrupación no supervisada de imágenes por atributos visuales.

Proyectos en curso

Además, mantenemos diversos trabajos de investigación activos con participación de estudiantes de pre y postgrado, y colaboradores académicos nacionales e internacionales. Aquí algunos de estos trabajos.

Unsupervised Learning for Sketch-Based Image Retrieval

Muchos de los modelos exitosos de visión por computador se basan en tener una gran cantidad de datos etiquetados. Sin embargo, en ambientes reales no es práctico etiquetar tal cantidad de datos. Así, con Javier Morales, memorista del Departamento de Ciencias de la Computación (DCC) de la Universidad de Chile, y Nils Murrugarra, investigador de Snap, estamos trabajando en métodos autosupervisados para el aprendizaje de representaciones visuales (*embeddings*) en el contexto de recuperación de imágenes. Además, apuntamos a crear modelos híbridos que aprendan a partir de datos etiquetados en forma supervisada y que al mismo tiempo se alimenten de datos no etiquetados para mejorar la generalización.

ColoSketch2Photo

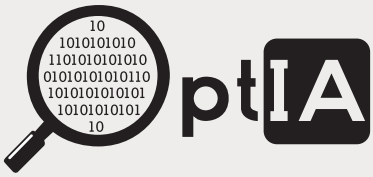
Convertir una expresión abstracta como lo es un dibujo a un objeto fotorrealista es de gran importancia en el *e-commerce*, especialmente en los rubros de personalización de productos. Los usuarios podrían dibujar lo que necesitan y obtener una representación real de esa abstracción. Junto a Diego Donoso, estudiante de magíster del DCC, estamos trabajando en diseñar modelos que permitan explotar la diversidad de dibujos que representan la semántica de una consulta y producir imágenes fotorealistas guiados por atributos adicionales como colores y texturas.

Invitación a colaborar

En Impresee eCommerce Labs buscamos producir conocimiento que permita mejorar el *e-commerce* tanto para los vendedores como para los mismos usuarios. Nos gusta colaborar con investigadores y formar equipos. Te invitamos a formar parte de estos y otros proyectos que ¡siempre tendrán un alcance nada menos que global! ▪

4 | https://impresee.com/scalable-visual-attribute-extraction/.

Iniciativas de Inteligencia Artificial



A continuación revisamos tres iniciativas nacionales recientes, gestadas con el objetivo de abordar la inteligencia artificial desde diversas perspectivas. Éstas incluyen el Instituto de Datos e Inteligencia Artificial (Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile), el Núcleo Inteligencia Artificial y Sociedad (Instituto de la Comunicación e Imagen, Universidad de Chile) y el Observatorio Público para la Transparencia e Inclusión Algorítmica (independiente).

Un Instituto de Datos e Inteligencia Artificial para Chile

FRANCISCO MARTÍNEZ Decano de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile.
MARCELA MUNIZAGA Directora Académica y de Investigación de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile.

El círculo virtuoso que se crea entre la disponibilidad masiva de datos y las herramientas que provee la inteligencia artificial ha sido identificado como la clave en una nueva etapa del desarrollo de la humanidad. Una etapa donde las capacidades humanas se expanden en una dimensión totalmente nueva, generando un nuevo espacio para la investigación científica y la tecnología de una magnitud difícil de imaginar. Desde hace algunas décadas venimos experimentando un cambio acelerado en todos los ámbitos de la sociedad traccionado por la fuerza de la revolución tecnológica que ha instalado una nueva red de infraestructura para la transmisión de datos a altas velocidades. Esta nueva y cambiante realidad permite automatizar muchas funciones de la vida, almacenar gigantescas bases de datos y explorar esos datos para generar información que se encuentra codificada en esas bases abriendo acceso a conocimientos científicos antes inexplorados. Estas nuevas capacidades son las que se exploran en la ciencia de los datos.

Se dice que los países que logren posicionarse como líderes en estos temas serán los que definan nuestro destino. Para algunos, los datos son lo que fueron las semillas, el oro, o luego el petróleo. Reconociendo la importancia de estos temas para el desarrollo de la ciencia y del país, en la Facultad de Ciencias Físicas y Matemáticas (FCFM) de la Universidad de Chile nos hemos planteado la pregunta de cómo abordar el desafío de contribuir en esta nueva ciencia. En esta reflexión hemos observado que los recursos vitales para la vida humana, como alimentos, minerales y la energía, los ha provisto la naturaleza a todo el planeta y la humanidad los ha transformado en bienes útiles a través de la historia tras procesos cada vez más complejos, hasta llegar a la revolución industrial. Con el tiempo esos procesos se han desarrollado con niveles crecientes de concentración de la producción hasta llegar a la actual globalización, que nos hace difícil participar del club de los grandes productores lo que nos relega al grupo de proveedores de recursos naturales. En el inicio de la era digital, en cambio, se perciben nuevas oportunidades para países como el nuestro de insertarse en la creación y producción de los bienes artificiales, cuya materia prima son los datos y cuyos productos que se generan utilizando un conjunto de algoritmos sofisticados muchos de ellos basados en inteligencia artificial.

En ese contexto, pensamos que nuestro país tiene potencial para convertirse en un actor relevante. Hemos desarrollado experiencia en manejo de grandes volúmenes de datos, como por ejemplo en el ámbito de la astronomía, y también en los sistemas que administran datos personales. Por ejemplo, el hecho de que cada persona al nacer o al llegar al país

reciba un número único que lo identifica, y que se utiliza para cualquier trámite que realice, genera un nivel de trazabilidad que no se da en otros países. Esto representa una enorme oportunidad para hacer análisis de esos datos, pero a su vez un desafío ético de cómo y para qué se usa esa información.

Las grandes preguntas de investigación van desde la teoría de la ciencia de los datos que busca identificar sus estructuras esenciales en grandes bases de información, hasta el diseño de algoritmos eficientes que se requieren para procesar y analizar los datos, pero también con las preguntas relacionadas con la ética que cuestionan el uso del poder asociado al control de la información. Otra observación a considerar sobre el asunto de cómo abordar el desafío de la Ciencia de los Datos es que en el expansivo universo de los datos concurren todas las disciplinas, como la astronomía, la biología, la sociología, la economía, la filosofía, entre otras. Es decir, la mirada desde los datos nos lleva a observar la naturaleza y la sociedad con ojos nuevos, de naturaleza digital que nos permiten ver aquello que hasta hace poco estaba en la oscuridad y soñar con explorar lo que permanece bajo el velo de la ignorancia.

En el caso de la FCFM, se cultivan varias disciplinas que tienen que ver con este tema. Hay investigación relacionada con ciencia de datos en los distintos departamentos como Ciencias de la Computación, Ingeniería Eléctrica, Ingeniería Industrial e Ingeniería Matemática. También hay centros de excelencia que lo abordan buscando conocimiento y soluciones a problemas concretos, como el Centro de Modelamiento Matemático (CMM), el Instituto Milenio Fundamentos de los Datos (IMFD), el Centro Avanzado de Tecnología para la Minería (AMTC) y el Instituto Sistemas Complejos de Ingeniería (ISCI). En general, todos los departamentos y centros de la FCFM utilizan datos y modelos para observar y predecir distintos fenómenos, como por ejemplo la astronomía, la observación y monitoreo del cambio climático, el monitoreo del comportamiento sísmico, entre otros. Un análisis nos mostró que la Universidad de Chile es la institución que más publicaciones ISI WoS tiene en el país en las áreas temáticas de Datos e Inteligencia Artificial.

Para abordar el cultivo de la Ciencia de Datos los países han hecho enormes inversiones, creando centros dedicados exclusivamente a ello, y muchas universidades en el mundo lo están abordando ya sea desde la estructura existente, o bien creando una nueva. En ese contexto, la FCFM decide crear una institucionalidad que permita desarrollar estos temas de forma inter y transdisciplinar, con la misión de agrupar y potenciar el trabajo que se realiza relacionado con ciencia de datos en las diferentes unidades y constituirse en un polo de pensamiento y creación en esta materia. Esta visión nos obliga a concebir una nueva institucionalidad capaz de permear las fronteras de departamentos y centros, y eventualmente también de facultades, generando un núcleo de investigación que concentre el aporte de las unidades e investigadores de diversos intereses científicos.

Con ese objetivo, se crea un Instituto de Facultad en Datos e Inteligencia Artificial, o ID&IA, que se proyecta como un centro referente a nivel nacional e internacional, con especial liderazgo en el ámbito latinoamericano. Esta iniciativa fue aprobada por el Consejo de Facultad en octubre de 2020 y ya ha dado sus primeros pasos, que consisten en la creación de un Comité Constituyente, liderado por el decano, con participación de 15 académicos de cinco departamentos, dos centros y dos institutos, y en la convocatoria a un concurso público para la contratación de tres nuevos académicos con dedicación exclusiva al Instituto. Además, el ID&IA se concibe con una lógica colaborativa inter y transdisciplinar, lo que se implementa permitiendo la doble adscripción, de manera que el claustro lo integren todos los académicos de los Departamentos de la Facultad con interés en el área de ciencia de datos y que los investigadores de los centros puedan también integrarse. Esta doble adscripción constituye una novedad en nuestra Facultad que permite que el Instituto sea efectivamente un núcleo atractor basado en la colaboración de todas las unidades de Beauchef.

De esta manera, el ID&IA podrá afrontar la misión de desarrollar las funciones académicas de investigación y transferencia de conocimiento, aportar en docencia de pregrado y postgrado y desarrollar extensión en las temáticas de datos e inteligencia artificial, atendiendo a los valores de la excelencia y el compromiso con la sociedad, de una manera multidisciplinar y promoviendo la colaboración entre los departamentos y centros, otras unidades académicas de la Universidad de Chile, y otras instituciones tanto nacionales como internacionales. Dentro de los objetivos del ID&IA se destaca el desarrollar investigación de alta calidad, apoyar la formación de académicos y profesionales de excelencia, contribuir al desarrollo nacional con soluciones innovadoras basadas en análisis de datos y en la construcción y aplicación de herramientas que utilizan inteligencia artificial, además de construir vínculos con otros centros nacionales e internacionales en las áreas de datos e inteligencia artificial.

Pensamos que el ID&IA, concebido con visión innovadora en su estructura y en la forma colaborativa de abordar las grandes preguntas, provee mejores condiciones para explorar la nueva dimensión del universo de los datos aunando las capacidades e intereses, fortalecidos con esa integración sinérgica para lograr enfrentar mayores desafíos y hacer contribuciones de mayor relevancia. Con esto, esperamos aportar significativamente al desarrollo sustentable del país y la región. La urgencia de este tema nos plantea un desafío que debemos abordar con mucho compromiso, poniendo todas nuestras capacidades al servicio de la comunidad. ▪

Núcleo Inteligencia Artificial y Sociedad [IA+SIC] Instituto de la Comunicación e Imagen

EQUIPO DIRECTIVO Ana María Castillo y Lionel Brossi, Instituto de la Comunicación e Imagen de la Universidad de Chile.

El Núcleo Inteligencia Artificial, Sociedad, Información y Comunicación IA+SIC[1] surge a partir de experiencias investigativas, de formación y de trabajo aplicado de sus integrantes, en ámbitos relacionados con el impacto que la inteligencia artificial tiene y tendrá en la sociedad, en áreas como la comunicación, la calidad de la información y el periodismo, la educación, el futuro del trabajo, entre otros.

La creación del Núcleo IA+SIC, se concibe en un escenario regional y de país, donde comienzan a desarrollarse iniciativas tendientes a diseñar políticas para la regulación, el desarrollo e implementación ética de sistemas de inteligencia artificial, que permean diversas áreas de la sociedad. Como ejemplo se encuentra la iniciativa gubernamental Política Nacional de Inteligencia Artificial y la Estrategia de Inteligencia Artificial propuesta por la Comisión Desafíos del Futuro, encabezada por el senador Guido Girardi.

IA+SIC conforma un núcleo interdisciplinario, creador de conocimiento y reflexión crítica a través de la investigación y desarrollo, que se ocupa además, de monitorear los desarrollos tecnológicos emergentes en el área de la inteligencia artificial y su implementación, con especial dedicación a los aspectos éticos, de gobernanza y consecuencias para el desarrollo social del país, desde un enfoque de respeto irrestricto a los derechos humanos, el pluralismo y la inclusión de diversidades.

El objetivo general del Núcleo es generar conocimiento científico y divulgación sobre los posibles impactos en términos de oportunidades y desafíos que implica el diseño, desarrollo e implementación de la inteligencia artificial en las personas, comunidades y en la sociedad en general, desde una mirada ética y de derechos humanos.

En lo específico, IA+SIC se propone desarrollar abordajes reflexivos y críticos en relación con la irrupción de tecnologías algorítmicas en la sociedad a partir de instancias investigativas, de formación, creación, de vinculación con el medio e internacionalización. El Núcleo promueve el diseño, desarrollo, implementación y utilización de la inteligencia artificial de manera que respete los valores sociales de equidad, diversidad y pluralismo con un enfoque de respeto a los derechos humanos.

1 | http://ia-sic.org.

A través de sus acciones, también apoya el desarrollo de iniciativas y políticas nacionales y regionales éticas sobre el diseño, desarrollo, implementación y uso de sistemas de inteligencia artificial en diversas áreas de la sociedad. Además, se propone la generación y consolidación de una comunidad local y nacional multisectorial (generadores de políticas públicas, academia, sociedad civil y sector privado) con foco en ética y gobernanza de la inteligencia artificial.

El objetivo específico dedicado a la incidencia en políticas públicas relacionadas al diseño, desarrollo, implementación y utilización de la inteligencia artificial ética e inclusiva, se manifiesta a través de las diferentes acciones, descritas a continuación.

En el año que lleva desde su creación, el Núcleo Inteligencia Artificial y Sociedad del Instituto de la Comunicación e Imagen, ha participado en numerosas iniciativas con impacto global, regional y nacional. Entre ellas, la participación para el diseño de las guías sobre inteligencia artificial y derechos de niñas, niños y jóvenes de UNICEF, a partir de la implementación de talleres participativos con jóvenes a lo largo de Chile, en las recomendaciones para generadores de políticas públicas de la International Telecommunications Union (ITU) volcadas en el reporte "Child Online Protection for policymakers", en las mesas de trabajo de las políticas de Inteligencia Artificial de Colombia y Perú. Para el caso chileno, el Núcleo ha colaborado en los esfuerzos para desarrollar la Estrategia Nacional de Inteligencia Artificial y es parte de la Subcomisión para la regulación de la ciberseguridad y de las plataformas digitales de la Comisión de Desafíos del Futuro, Ciencia, Tecnología e Innovación del Senado.

Desde el año 2020, el Núcleo participa de la Mesa para el desarrollo de la estrategia nacional contra la desinformación del Consejo para la Transparencia, específicamente coordinando la submesa encargada del diseño e implementación del plan de formación y difusión, previsto para 2021.

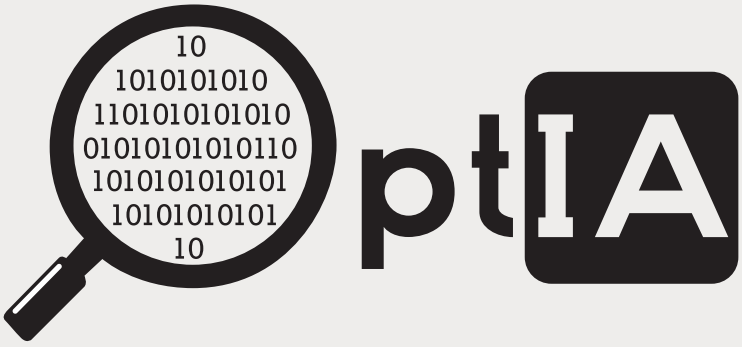
Asimismo y en conjunto con la fundación Wikimedia Chile, se lanzó el Webcast Utopías dentro y fuera de las pantallas, donde participaron líderes de diversas organizaciones globales para discutir temas como la educación, los derechos, la ética, los datos abiertos, entre otros, en su relación con el campo de la Inteligencia Artificial.

Entre los proyectos de investigación actualmente vinculados al Núcleo IA+SIC se encuentran "Hablatam[2]: Jóvenes, habilidades digitales, brechas de contenido y calidad de la información en América Latina", financiado por la Agencia Nacional de Investigación de Uruguay y la Fundación Ceibal a través del Fondo Sectorial de Educación, Modalidad Inclusión Digital; el proyecto "Future Ways of Working in the Digital Economy"[3], financiado por la Agencia Nacional de Investigación de Noruega; el proyecto "Jóvenes, medios digitales y discursos públicos de pandemia en América Latina", desarrollado en conjunto con el Centro Heidelberg para América Latina y el Núcleo Milenio IMHAY, y el proyecto "Desafíos éticos para la docencia de pregrado en el contexto del desarrollo e implementación de sistemas de inteligencia artificial en la educación", financiado por el Departamento de Pregrado de la Vicerrectoría de Asuntos Académicos de la Universidad de Chile. ▪

2 | http://conectadosalsur.org/hablatam.

3 | https://www.bi.edu/research/centres-groups-and-other-initiatives/futurewaysofwork/.

OptIA: Observatorio Público para la Transparencia e Inclusión Algorítmica



DIRECTORIO OPTIA Ricardo Baeza-Yates, Alejandro Barros, Daniel Vak Contreras, Carol Hullin, Óscar López, Catherine Muñoz, Claudia Negri, Luis Pizarro y Danielle Zaror.

Somos testigos de un periodo sin igual en la historia de la humanidad. Gran parte de nuestra vida personal, nuestra convivencia en la sociedad y la comprensión del mundo que nos rodea está siendo mediada por la tecnología a niveles que las personas no imaginan.

En medio del vendaval de decisiones automatizadas y los procesos que éstas desatan, encontramos una sociedad que apenas tiene capacidad de reacción y mucho menos idea sobre cómo regular los fenómenos y consecuencias de esta vorágine tecnológica.

Chile no es la excepción; nuestro país no cuenta con regulaciones apropiadas en materia de protección de datos, de ciberseguridad ni de delitos informáticos. Recientemente se ha comenzado a discutir una política nacional de inteligencia artificial que omite los déficits anteriores y ni siquiera contempla una gobernanza ni recursos para hacer frente a los desafíos que una tecnología como ésa supone, y que ya es aplicada por empresas del sector privado y, lo que resulta más preocupante aún, también en el sector público.

Este escenario fue el que motivó a un conjunto de profesionales, de diversos orígenes y disciplinas, quienes durante la pandemia nos convocamos de modo virtual para conversar sobre nuestras inquietudes, para finalmente embarcarnos en la tarea de crear un Observatorio para la Transparencia y la Inclusión Algorítmica. Es por esto que desde OptIA nos hemos propuesto aportar desde una mirada profesional y multidisciplinaria sobre estas temáticas.

Nos preocupa principalmente, pero no exclusivamente, la implementación de soluciones tecnológicas de inteligencia artificial adoptadas por el Estado. Muchas de estas iniciativas se presentan como infalibles y prometen mejorar ciertos procesos y tomas de decisiones sin mayor transparencia en su funcionamiento y su alcance. Se trata además de sistemas que no tienen declarado un control sobre su impacto en la sociedad, en la privacidad ni el tratamiento de los datos que utiliza, y que pueden (ciertamente) profundizar los sesgos, la discriminación y la asimetría de poder cuando dichos sistemas toman decisiones injustas.

En OptIA compartimos la preocupación sobre la afectación de grupos vulnerables, históricamente marginados y excluidos, compartiendo asimismo la necesidad de ser un agente colectivo de cambio para la generación de políticas públicas justas e inclusivas en relación con estas tecnologías.

La implementación de la estrategia nacional de inteligencia artificial no ha sido suficientemente discutida, y por lo tanto creemos que la implementación de una política pública en un tema tan relevante para los próximos años debe tener un proceso de discusión y de participación amplio con todos los sectores del país.

La falta de representatividad de la sociedad civil en las discusiones y toma de decisiones relacionadas a las tecnologías digitales emergentes y aquellas que usan algoritmos y/o inteligencia artificial, hacían urgente el surgimiento de organizaciones como la que hemos levantado. Nuestro objetivo es velar porque la práctica tecnológica considere la elaboración de algoritmos inclusivos, que consideren la diversidad de la sociedad, y que respondan a requerimientos basados en los derechos humanos.

Los sistemas de inteligencia artificial (IA) utilizados en políticas públicas han demostrado, según abundante evidencia internacional, fallar continuamente en temas tan delicados como vigilancia policial predictiva, análisis predictivo de bienestar infantil, evaluación de riesgos y los sistemas de decisión de beneficios públicos, por lo que es necesario, en base a una política de riesgos, contar con prácticas vinculantes específicas, que incluyan al menos las siguientes consideraciones:

- Los organismos públicos no deben adquirir ni utilizar sistemas que estén protegidos de revisión pública, tales como secretos industriales o acuerdos de confidencialidad.
- Debe existir transparencia activa, no a petición de parte, con mecanismos como registro de algoritmos y plataformas disponibles al público.
- Evaluaciones de impacto algorítmico que analicen tanto los riesgos como los beneficios que supone tener un determinado sistema, elaboradas por terceros expertos e independientes.
- Debe existir personal capacitado para la implementación, uso y mitigación de sistemas de IA.
- Procesos de licitación competitivos y abiertos.
- La colaboración público-privada debe ser totalmente transparente, haciendo público conflictos de intereses, contratos con proveedores y cualquier información relevante, cumpliendo con las más altas exigencias de probidad y rendición de cuentas.
- Se debe evaluar la afectación de las personas más vulnerables y la posibilidad que éstas puedan hacer sus propias evaluaciones y oponerse a determinadas implementaciones.
- Se debe evaluar si el sistema de IA crea las condiciones y la capacidad para supervisión humana significativa, que incluye la supervisión de aquellos que se ven directamente afectados por estos sistemas.

Como sabemos que lograr marcos regulatorios en materias como éstas son desafíos gigantescos, en OptIA trabajaremos y promoveremos el reconocimiento de al menos los siguientes principios para la implementación de soluciones automatizadas y de inteligencia artificial con el fin de proveer herramientas éticas para la resolución de los conflictos que sabemos se presentarán:

Proporcionalidad e inocuidad: en su virtud, promoveremos que se elija un método de inteligencia artificial cuando esté justificado y sus resultados sean convenientes para los fines perseguidos una vez aplicadas evaluaciones de costo versus beneficio. Un método será inocuo cuando su aplicación no genere daños a los seres humanos, al medio ambiente y a los ecosistemas.

Inclusión y no-discriminación: la inteligencia artificial debe ser un mecanismo que genere justicia social de manera que sus beneficios deben buscarse procurando alcanzar al mayor número de personas posible sin distinción de etnia, edad, situación migratoria, identidad de género o nivel socioeconómico. Cada vez que se produzca un resultado discriminatorio, los administradores de la tecnología de inteligencia artificial deben incluir mecanismos para apelar ese resultado, debiendo revisarse las características de los algoritmos utilizados y sus bases de datos.

Transparencia y explicabilidad: las personas tienen derecho a saber cuándo se toma una decisión sobre la base de algoritmos y, en esas circunstancias, exigir o solicitar explicaciones e información a empresas del sector privado o instituciones del sector público.

Privacidad y seguridad: se trata de una garantía fundamental que debe cautelarse durante todo el ciclo de vida de los sistemas de inteligencia artificial, debiendo establecerse marcos de protección y mecanismos de gobernanza adecuados, respaldados por los sistemas judiciales en caso de infracción.

Autonomía y supervisión humana: el ser humano siempre debe poder autodeterminarse, de manera que conserve el poder de decidir qué decisión tomar sobre sí mismo, en lugar de que lo haga un sistema de IA. Siempre debe ser posible atribuir la responsabilidad ética y jurídica, en cualquier etapa del ciclo de vida de los sistemas de IA, a personas físicas o a entidades jurídicas existentes. Esta supervisión hu-

mana no es sólo individual, sino que también se refiere a la supervisión pública dentro de la que se insertan organizaciones no gubernamentales como OptIA.

Responsabilidad y rendición de cuentas: los creadores de sistemas de inteligencia artificial deben asumir las consecuencias éticas y jurídicas de las tecnologías que diseñen e implementen de conformidad con el ordenamiento jurídico vigente. La obligación de rendir cuentas debe sustentarse en mecanismos adecuados de supervisión a lo largo de todas las etapas, para esto la auditabilidad y trazabilidad de los procesos son una condición esencial.

Nuestras acciones y el futuro cercano

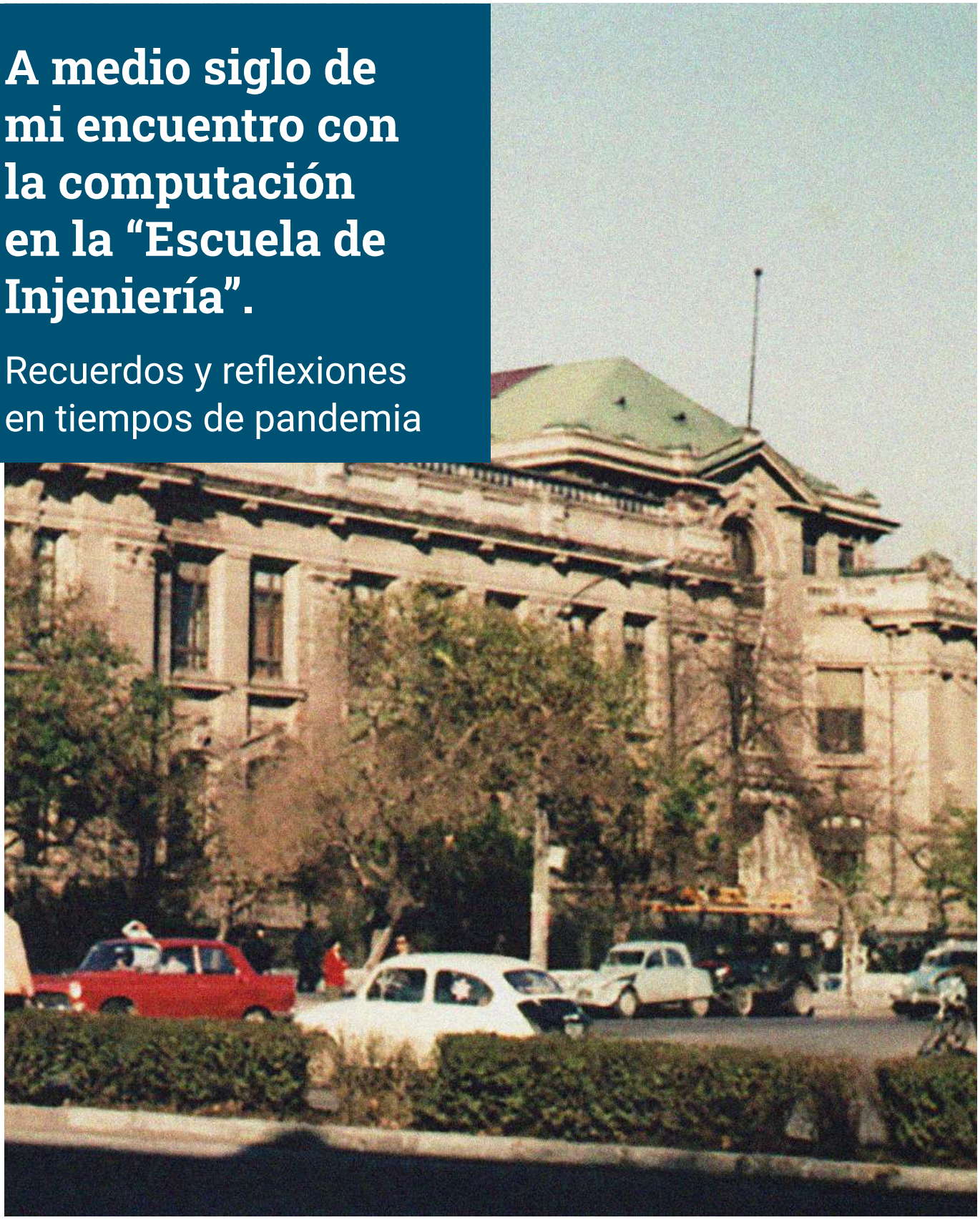
Una de nuestras primeras acciones fue participar de la consulta pública sobre la Política Nacional de Inteligencia Artificial de Chile. Así elaboramos un documento[1] con nuestros comentarios y recomendaciones en respuesta a la referida consulta. Algunas de nuestras recomendaciones apuntaron a cambiar la definición de IA para efectos regulatorios y políticas públicas, ya que es necesario que la definición se centre no sólo en el componente técnico (IA estrecha), sino también en las estructuras sociales que la rodean y en los impactos sobre las personas, especialmente aquellos más vulnerables, y en la importancia por el respeto a la dignidad humana. Una definición netamente técnica puede llevar al sesgo de automatización o ignorar los impactos sociales que son un problema real a nivel global.

También hemos sido parte de la organización del XII Encuentro Internacional de IA en enero de 2021 junto al Instituto Milenio Fundamentos de los Datos, donde participaron nuestros directores Ricardo Baeza-Yates y Catherine Muñoz, y durante abril de 2021 nuestras directoras Claudia Negri y Danielle Zaror fueron parte de la conversación sobre la Agenda Digital para la nueva Constitución.

1 | https://optia.cl/2021/01/29/respuesta-a-la-consulta-sobre-politica-nacional-de-ia/.

A medio siglo de mi encuentro con la computación en la “Escuela de Injeniería”.

Recuerdos y reflexiones en tiempos de pandemia



JUAN ÁLVAREZ RUBIO

Académico del Departamento de Ciencias de la Computación de la Universidad de Chile. Master of Mathematics (Computer Science), University of Waterloo. Ingeniero de Ejecución en Procesamiento de la Información, Universidad de Chile. Junto a su labor como docente, trabaja en reconstruir la historia dela computación en Chile.

jalvarez@dcc.uchile.cl

La motivación por escribir este artículo surgió de una entrevista acerca de los inicios de la computación en la Universidad de Chile que nos hizo Patricio Aceituno, ex decano de la Facultad de Ciencias Físicas y Matemáticas, en una semana de actividades dedicadas a la innovación en noviembre de 2020. Los temas que surgieron de la conversación junto a mi compañero del Centro de Computación, Julio Zúñiga, y el estado reflexivo en que nos tiene esta interminable pandemia, me alentaron a escribir por primera vez en primera persona acerca de la historia de la computación en la "Escuela de Injeniería" de la Universidad de Chile. Si bien casi todos los temas los he desarrollado más formalmente en otros artículos, los cincuenta años de mi llegada a la Facultad, que cumplí en marzo del año 2021, me decidieron a escribir mi experiencia personal con la computación.

Introducción

Ingresé a la Escuela de Ingeniería de la Universidad de Chile en 1971, un año especial y de mucha esperanza en el país, donde se ampliaron significativamente las vacantes para que ingresáramos también estudiantes de los sectores obreros y populares, que habíamos estudiado en escuelas y liceos fiscales desde donde egresamos en 1970 de la última generación de sexto año de la educación secundaria. Quedé seleccionado tanto en la U como en la UC, ambas gratuitas. Mi conocimiento de computación se limitaba a saber que las tarjetas que utilizamos para marcar las respuestas de las pruebas de selección eran procesadas por un computador y que la Prueba de Aptitud Académica se había rendido por primera vez en enero de 1967, reemplazando al anterior y cuestionado Bachillerato.

Decidí matricularme en la Universidad de Chile porque me sentí más cómodo por coincidir con mis compañeros de un liceo público no emblemático y de otros liceos fiscales, que desde entonces preferían a la Universidad de Chile. Llegamos al imponente edificio de la "Escuela de Injeniería" ubicado en Beauchef 850. El viejo recinto, que ya tenía cerca de medio siglo, albergaba una moderna Facultad de Ciencias Físicas y Matemáticas que se había reestructurado en departamentos en 1964 y que en 1970 había renovado su sistema docente anual y rígido a uno semestral y flexible. El moderno sistema curricular fue mi segundo contacto con la computación. A través de tarjetas perforadas se registraban nuestras preferencias por cursos y profesores y el sistema nos inscribía en las distintas secciones de las asignaturas. Elegí mis cursos de álgebra y cálculo con el notable y legendario profesor Moisés Mellado, una de las secciones del laboratorio de física, y un curso de computación.

El primer curso de computación

La asignatura "Introducción a la Computación" tenía el código MA151 porque la impartía el Departamento de Matemáticas que había sido creado en 1964 y contaba con un grupo de investigadores en Computación. Mi profesor fue Víctor Canales, un joven ingeniero matemático que trabajaba en ECOM, la empresa nacional y estatal de computación, que había sido creada en 1968 para dar servicio a las instituciones públicas y entrenar a programadores, analistas de sistemas y operadores que se necesitaban para los enormes, costosos y muy escasos computadores de la época.

En la primera parte del curso se estudiaba la estructura de los computadores incluyendo la representación binaria de instrucciones y datos utilizando los apuntes de "Introducción a la Computación" del profesor Víctor Sánchez. La segunda parte era una introducción a la programación basada en los apuntes del profesor Pablo Fritis y de los ayudantes de investigación Félix Aguilera y Fernando Gamboa. El lenguaje de programación era FORTRAN (FORmula TRANslator), el lenguaje emblemático para aplicaciones "científicas" y de "ingeniería". Para facilitar el proceso de los programas se utilizaba WATFOR (WATerlooFORtran), un software construido en la Universidad de Waterloo de Canadá que permitía ejecutar un lote de varios programas en FORTRAN de una manera mucho más eficiente. Por otra parte, una de las seis secciones del curso utilizaba ALGOL (ALGOrithmic Language) y estaba a cargo del profesor Herbert Plett (ingeniero eléctrico e investigador del área de computación del Departamento de Matemáticas).

Entonces se programaba según el paradigma imperativo en que las instrucciones de control básicas eran *if* (sin

else) y *goto* (que saltaba o bifurcaba a una instrucción que no era la siguiente en la secuencia). Los programas resultaban desordenados y difíciles de comprender. De hecho, antes de programar se debía expresar el algoritmo de solución dibujando un diagrama de flujo, que era una representación gráfica de la lógica o flujo de control de la ejecución de las instrucciones para lo cual existían unas regletas para dibujar las formas estandarizadas de representación de las distintas instrucciones. Una vez elaborado el diagrama de flujo, sus elementos se traducían en instrucciones del lenguaje y se escribían en papel o en "hojas de codificación" de 24 líneas de 80 caracteres (usando sólo letras mayúsculas, dígitos y algunos pocos signos especiales).

Una vez codificadas las instrucciones del programa en FORTRAN se debían perforar en tarjetas de 80 columnas (ver Figura 1), para lo cual era necesario conseguir una de las pocas máquinas perforadoras que existían, y entre ellas ojalá una KP-29 en lugar de las más antiguas y básicas KP-26. Después de esperar que se desocupara alguna máquina, estaba la tarea de perforar las tarjetas. Y no había que equivocarse porque las perforaciones incorrectas inutilizaban la tarjeta y había que reemplazarla. Y había que cuidar que no se "cayera el sistema", es decir, el mazo de tarjetas y se desordenaran las instrucciones.

Las tarjetas se entregaban en una oficina que prometía, en el mejor de los casos, una respuesta al día siguiente. La desilusión se producía al recibir las tarjetas con un listado impreso señalando que se habían detectado errores de sintaxis. Por lo tanto, había que regresar a las máquinas perforadoras para rehacer las tarjetas incorrectas. Después de un par de días, y una vez corregidos todos los errores de sintaxis, aparecían los errores de ejecución, es decir los resultados incorrectos. De vuelta a corregir e iterar. Todas estas etapas, que en el mejor de los casos tardaba una semana, los estudiantes de hoy lo logran en algunas horas en sus computadores personales.

En resumen, los más afortunados conseguíamos los resultados correctos sin tener acceso al computador IBM/360 (llamado así porque pretendía abarcar los 360 grados de todo el espectro de aplicaciones) y al que sólo podíamos contemplar extasiados detrás de una vidriera, asombrándonos del parpadeo de las luces del panel de control, de los movimientos de las unidades de cintas magnéticas y de la lectura vertiginosa de las tarjetas. Era un enorme computador, el más grande en Latinoamérica,

A pesar de toda la burocracia para poder usar el computador, pero sin interactuar directamente con el IBM/360, algunos fuimos seducidos/abducidos por el entonces "arte" de programar y nos decidimos estudiar esa especialidad.

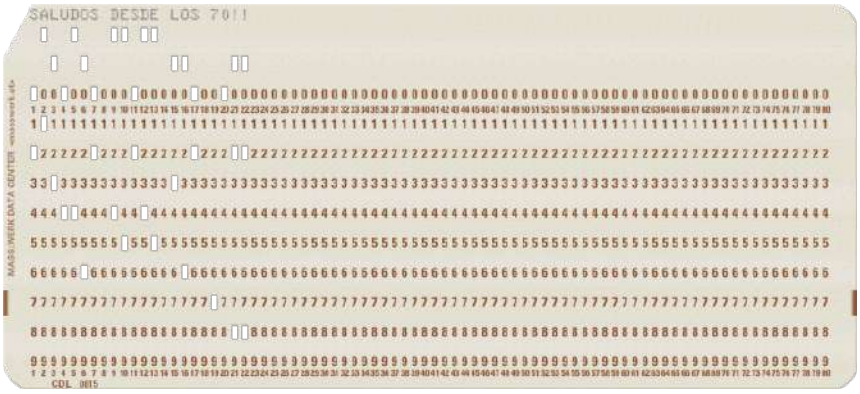


Figura 1. Tarjeta de 80 columnas para perforar programas y datos para el computador IBM/360.

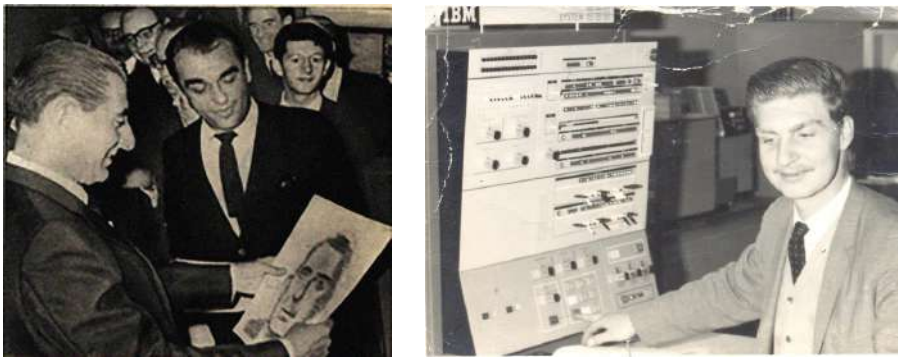


Figura 2. Eduardo Frei, Efraín Friedman y un operador del computador IBM/360. Año 1967.

Figura 3. Pablo Fritis, Hugo Segovia y Víctor Sánchez, creadores de las carreras de Computación. Año 2009.

y tuvo un costo cercano al millón de dólares y fue inaugurado por el presidente Eduardo Frei Montalva en enero de 1967 (ver Figura 2). Tenía 128Kb de memoria y residía en una enorme sala en el subterráneo de la torre central con un piso "falso" para el cableado eléctrico que conectaba las distintas unidades (unidad central de proceso, consola de operación, lectora de tarjetas, impresora, 4 unidades de cinta y 2 unidades de disco) y un techo "falso" con los equipos de aire acondicionado para mantener la temperatura adecuada para su funcionamiento.

La carrera de Computación

A pesar de toda la burocracia para poder usar el computador, pero sin interactuar directamente con el IBM/360, algunos fuimos seducidos/abducidos por el entonces "arte" de programar y nos decidimos estudiar esa especialidad. Después de dos semestres de Plan Común, se podía ingresar a la recién creada carrera de Ingeniería de Ejecución en Procesamiento de la Información (IEPI). En lo personal me vino muy bien ingresar a una carrera de ocho semestres que me permitiría aliviar pronto la carga económica a mi modesta familia.

La IEPI fue la primera carrera de ingeniería en el área de computación en el país y fue la sucesora de la carrera de Programación de Computadores de tres años de duración creada en 1968. Los planes de estudios fueron diseñados, sin disponer entonces de referentes internacionales, por tres investigadores del área de Computación del Departamento de Matemáticas: Héctor Hugo Segovia, ingeniero industrial; Pablo Fritis, ingeniero civil; y, Víctor Sánchez, ingeniero industrial mecánico de la Universidad Técnica del Estado (ver Figura 3). En los primeros años de los setenta, Segovia y Fritis asumieron responsabilidades directivas en ECOM y Sánchez se trasladó a la Universidad Técnica, asumió la dirección de su Centro de Computación y creó la carrera de Ingeniería de Ejecución en Computación e Informática y sus alumnos procesaban sus trabajos en el IBM/360 de la Universidad de Chile.

Cabe señalar que en diciembre de 1964 el decano Enrique D'Etigny había presentado en el Consejo Universitario una propuesta de carrera de Ingeniería en Computación que fue rechazada por no entenderse aún la necesidad de esos profesionales cuando en el país existía sólo una decena de computadores. En enero de 1965 se aprobó en cambio una carrera de Ingeniería Matemática con una orientación distinta y de cinco años de duración.

En primer lugar, inscribimos el curso de "Estructuras y procesos de información" con el profesor Pablo Fritis. Era un curso de estructuras de datos y algoritmos y se programaba en el lenguaje PL/I, un lenguaje diseñado por IBM para desarrollar aplicaciones, tanto científicas y de ingeniería, como "comerciales" o administrativas. El lenguaje permitía la programación estructurada, que mejoró el estilo "spaghetti" de la programación imperativa, al disponer de las instrucciones *if-else* y *while*. Además, PL/I tenía facilidades para manejar archivos secuenciales y de acceso directo (por posición relativa o por llave).

Paralelamente, cursamos "Programación de Computadores I" con el profesor Víctor Sánchez, único curso de 13 Unidades Docentes, con 3 clases semanales y una clase auxiliar de 2 horas. Se programaba en el lenguaje Assembler/360, notación simbólica del lenguaje binario de máquina del computador IBM/360 siguiendo el "Manual de Assembler" del profesor Sánchez. Junto con el curso siguiente, "Programación de Computadores II" (que cursé con el profesor Sergio Gamboa), con énfasis en uso y construcción de macroinstrucciones, las dos asignaturas proporcionaban una introducción a la programación de sistemas, es decir, al desarrollo de programas "utilitarios" complementarios del sistema operativo o de aplicaciones críticas que requerían un uso eficiente de los escasos recursos disponibles de memoria y tiempo de procesador. Más adelante, el curso de "Sistemas Operativos" también se orientó al sistema operativo del computador IBM/360, y el profesor fue un ingeniero de la IBM.

El resto de los cursos obligatorios de especialidad incluía asignaturas de: Tecnología de Equipos (sobre las máquinas Hollerith o Unit-Record, equipos especializados en diferentes procesos *off-line* con las tarjetas), Programas de Aplicación (especialmente para control de proyectos y programación lineal e investigación operativa), Técnicas de Procesamiento de Datos (orientadas al desarrollo de sistemas de información administrativos y de su programación en el lenguaje COBOL) y un Taller de Procesamiento de Datos (con el desarrollo de un proyecto de mayor envergadura durante todo un semestre).

Mención especial merece el curso de "Lenguajes y Compiladores" con los jóvenes profesores Fernando Gamboa y su auxiliar Patricio Poblete, ambos del grupo de computación del Departamento de Matemáticas. Después de la introducción sobre autómatas y lenguajes formales, desarrollamos analizadores léxicos y compiladores. El curso me fascinó, al punto que al año siguiente fui profesor auxiliar del recién asumido profesor de cátedra Patricio Poblete.

Entre los cursos electivos recuerdo especialmente el de Sistemas de Información con el profesor Hugo Segovia y el de Simulación con el joven ingeniero Hernán Avilés. Ambos trabajaban en ECOM y estuvieron involucrados en el desarrollo del emblemático proyecto Synco o Cybersyn, que se desarrolló entre los años 1971 y 1973 y cuyo propósito fue coordinar y planificar la producción en las empresas del área de propiedad social. Simulamos sistemas usando GPSS y Dynamo, la herramienta que se estaba usando para simulación dinámica en la componente CHECO, del sistema Synco, cuyo propósito era el desarrollo y planificación del aparato industrial.

Los cursos obligatorios incluyeron también Cálculo Numérico, Estadística, y Complementos de Matemáticas. Del Departamento de Industrias, los cursos de Introducción a la Economía y de Administración de Empresas. Y del desaparecido Departamento de Estudios Humanísticos, los cursos obligatorios de Filosofía, Ciencias Sociales e Inglés (general y especializado para computación). Por mi cuenta inscribí los cursos libres de Ciencia Política (con el profesor Jaime Castillo Velasco), Sociología (con el profesor Cumsille) e Historia de Chile (con la joven historiadora María Angélica Illanes).

Figura 4. De izquierda a derecha, de pie: Alfredo Piquer, Eugenio Bravo, Víctor Salas, J. Ricardo Giadach, Martín Borack, Julio Zúñiga, Pedro Vergara. Sentados: Claudio Vergara, Jaime De Mayo, Rafael Hernández, Marcelo Energici, Osvaldo Schaerer, Juan Álvarez. Conmemoración 40 años de contrato en CEC. Año 2013.

Mi educación formal terminó en los cuatro años de duración de la carrera. Mis profesores fueron ingenieros de otras especialidades que fueron parte de los pioneros de la computación en Chile. Trabajaban en empresas e instituciones del Estado (ECOM, Endesa, U, UTE, etc.) y, por lo tanto, además de los conocimientos técnicos, nos transmitieron una profunda vocación de servicio público. El agitado contexto sociopolítico de esos años fueron el telón de fondo de nuestra formación que nos estimuló y nos hizo tomar aún más conciencia del aporte individual y colectivo que podríamos hacer al país como parte de las primeras generaciones de una nueva y pujante disciplina de ingeniería.

El Centro de Computación

En 1972, cursando el segundo año de mi carrera, apareció en un fichero un aviso para concursar a cargos de ayudantes de investigación para el Centro de Computación (CEC). El concurso estaba abierto a todos los estudiantes de la Facultad y el único requisito era tener aprobado el curso de Introducción a la Computación. Recuerdo que se presentaron muchos postulantes y quedé seleccionado junto a Margarita Sprovera, Marcelo Energici, Rafael Hernández, Miguel Pérez, Jaime De Mayo, Claudio Vergara, Juan Carlos Rojas y Osvaldo Schaerer. Sólo yo era estudiante de IEPI, pero finalmente seis de nosotros nos titulamos de esa carrera.

Figura 5. Fernando Silva, Patricio Poblete, Alfredo Piquer, Nancy Hitschfeld. Conmemoración 35 años del DCC. Año 2010.

Fuimos contratados en octubre de 1972 y recibidos por Fernando Silva, Director del CEC y Carlos Pérez, encargado del grupo de Extensión. Inicialmente fuimos ayudantes de nuestros tutores Julio Zúñiga, Alfredo Piquer, Ricardo Giadach, Pedro Vergara, Víctor Salas y Gerardo Kahn, que eran también estudiantes de ingeniería matemática y eléctrica que habían ingresado al CEC un par de años antes que nosotros (ver Figura 4). Nos asignaron una oficina común con "sillas calientes" que ocupábamos entre nuestras clases. El despacho estaba en el subterráneo del edificio de Química, donde vivía sus últimos días el "Lorenzo" (el Standard Electrik Lorenz ER-56), un computador transistorizado que llegó en junio de 1962 y que fue el primer computador universitario en Chile y el tercero en el país.

Prontamente fuimos incorporados a los distintos proyectos de los dos grupos del CEC. El grupo de Extensión desarrollaba principalmente sistemas computacionales para usuarios universitarios: proyectos para distintas facultades y sedes de la universidad y los sistemas centrales de selección de alumnos, matrícula y administración docente. Por otra parte, el grupo de sistemas desarrollaba software incrustado o complementario a los sistemas operativos. Tuve la oportunidad de trabajar en proyectos de los dos grupos: sistema de selección de alumnos y programación de sistemas. Y en este nuevo trabajo pude por fin tener acceso directo al computador IBM/360 y financiar todos mis gastos, logrando este objetivo desde abril de 1973, fecha de mi primer sueldo.

El año 1973 fue muy especial. Permanecía todo el día en la Escuela y fui testigo directo de lo bueno y lo malo que sucedía. El día 11 de septiembre como siempre llegué muy temprano, entonces vivía en Renca y los que viven lejos siempre llegábamos antes. Me enteré del golpe porque me extrañó que no llegara ninguno de mis compañeros de oficina. Permanecí "defendiendo" la Escuela hasta el mediodía y regresé a trabajar el primer día que se reabrió. Este episodio es más largo y merecería otro artículo.

Parte importante del trabajo en el CEC era participar y desarrollar actividades y cursos de capacitación y difusión. Paralelamente, en la medida que nuestros tutores fueron haciéndose cargo de los cursos de IEPI como profesores, nosotros trabajamos primero como ayudantes, después como profesores auxiliares, y, a nuestro egreso, como profesores. Adicionalmente, colaboramos como profesor *ad honorem* en la Universidad Técnica del Estado (hoy USACH) en la carrera de Ingeniería de Ejecución en Computación e Informática y en el Instituto Politécnico de la Universidad de Chile (hoy UTEM) en la carrera de Programación.

A fines del año 1974 se decretó una rebaja temporal y sustancial de las tasas de importación de computadores. Consecuentemente, se produjo un ingreso masivo de máquinas para las cuales no existían entonces suficientes especialistas. Para aminorar el déficit, el CEC, junto con ECOM y la Asociación de Centros Universitarios de Computación, organizaron un Plan Nacional de Capacitación Intensiva en Computación (PLANACAP) para capacitar analistas, programadores y operadores. Participé en cursos para formar analistas y programadores, y, en mi calidad de programador de sistemas del CEC, dicté cursos de Assembler/360 para CODELCO en Antofagasta y en Rancagua.

El Departamento de Ciencias de la Computación

El Departamento de Ciencias de la Computación (DCC) fue creado el 1 de enero de 1975. Su primer director fue Fernando Silva (que además era director del CEC) y sus académicos fundadores fueron los ingenieros matemáticos José Pino, Alfredo Piquer y Patricio Poblete (ver Figura 5), el ingeniero civil electricista Miguel Guzmán, el químico y magíster en Ingeniería Eléctrica

Francisco Oyarzún, y los ayudantes de investigación Rafael Hernández y Patricio Zúñiga, ambos estudiantes de computación. A diferencia del CEC, que era un centro de servicio para toda la Universidad, el DCC era un departamento académico de la Facultad de Ciencias Físicas y Matemáticas con funciones de docencia, investigación y extensión.

En su primer año, el DCC propuso infructuosamente crear una carrera de Ingeniería Civil en Computación. Sí logró que la Facultad aprobara un Bachiller (de 4 años) y un Magíster (de 6 años) en Ciencias con mención en Computación. Por otra parte, heredó la carrera de IEPI que entonces tenía alrededor de 100 estudiantes y que fue creciendo año a año hasta alcanzar 400 alumnos en 1983 (llegando a ser la segunda en cantidad de alumnos en la Facultad), año en que se creó la Licenciatura en Ciencias de la Ingeniería (de 4 años) y la carrera de Ingeniería Civil en Computación (de 6 años).

En lo personal continué haciendo clases en los cursos de Computación en el Plan Común donde impulsamos, junto a otros colegas, cambios metodológicos y de paradigmas de programación y de lenguajes (ALGOL-W, RATFOR, Pascal, Turing, Java y Python). Por otra parte, y dada mi experiencia como programador de sistemas en el CEC, fui profesor durante muchos años del curso de "Programación de Computadores I" de IEPI y comencé a dictar el nuevo curso de "Programación en Lenguajes orientados a la Máquina" (PLOM) del Bachiller y de la Licenciatura. En este último, además de Assembler, se programó en PL360, un lenguaje estructurado para escribir programas para el IBM/360, y posteriormente en C, un lenguaje para programación de sistemas independiente de la arquitectura del computador. Algunos semestres se utilizó también el lenguaje Assembler de la arquitectura de un computador VAX, que fue prestado por la empresa SONDA, representante en Chile del fabricante Digital.

El agitado contexto sociopolítico de esos años fueron el telón de fondo de nuestra formación que nos estimuló y nos hizo tomar aún más conciencia del aporte individual y colectivo que podríamos hacer al país como parte de las primeras generaciones de una nueva y pujante disciplina de ingeniería.

Figura 6. Izquierda: Edificio Blanco Encalada 2120. Derecha: Julio Zúñiga, José M. Montecinos, Claudio Vergara en computador IBM/370.

En 1975 la Universidad adquirió un computador IBM/370 que se instaló en el segundo piso del recién inaugurado edificio "de Computación" ubicado en Blanco Encalada 2120 (ver Figura 6). El IBM/370 modelo 145, que costó un millón y medio de dólares, tenía 1 Mega de Memoria, 3 discos 3330 de 100Mb y 3 discos 3340 de 70Mb, 6 unidades de cinta, 2 impresoras, 2 lectoras de tarjetas, 16 terminales 2741 y 4 estaciones de despliegue 3277. El sistema operativo VM/370 simulaba máquinas virtuales que podían correr los sistemas operativos CMS, OS/VS1 o DOS/VS. Como programador de sistemas del CEC me correspondió programar extensiones al sistema operativo OS/VS1 para controlar y medir el uso de los recursos computacionales.

El IBM/370 representó un salto tecnológico cualitativo que facilitó la docencia y la investigación en la disciplina y se le recuerda especialmente por la introducción de las pantallas y los terminales distribuidos, un lustro antes de la aparición y rápida difusión de los computadores personales.

En julio de 1979 fui contratado como académico de jornada completa del DCC, lo que paradojalmente me significó bajar algunos grados en la escala única de sueldos, justo el mes en que contraje matrimonio. Además del aumento en mis responsabilidades docentes, tuve a mi cargo un computador Burroughs 1900 que fue cedido al DCC y participé en el proyecto de desarrollo de un software de recuperación de información (BIRDS) bajo la guía de José Pino, director del DCC, y que diseñó el sistema junto a Alfredo Piquer y Patricio Poblete. Adicionalmente, en 1979 José Pino creó y fue director de la revista Informática, donde tuvimos oportunidad de escribir varios artículos de difusión especialmente dirigidos a los programadores de la industria.

Figura 7. Miguel Johnatan (UFRJ), José Pino, M. Cecilia Rivara, Ignacio Casas (UC), Juan Álvarez. I Congreso Iberoamericano de Educación Superior en Computación. Año 1991.

Figura 8. Claudio Gutiérrez, Aldo Migliaro. José Acle, Isaquino Benadof, Guillermo González, Víctor Sánchez, Wolfgang Riesenkönig, Juan Álvarez. I Taller de Historia de la Computación en Chile. Año 2009.

A fines de los setenta y comienzos de los ochenta éramos muy pocos académicos de tiempo completo. Recuerdo algunos años en que habíamos sólo tres o cuatro en el DCC, mientras otros estaban en sus posgrados. Nos correspondió por tanto dictar diversos cursos contando con la ayuda de ingenieros del CEC y de empresas y algunos de los primeros egresados de IEPI. Mi primera etapa en el DCC culmina en 1983 con mi viaje a la Universidad de Waterloo en Canadá y mi obtención del grado de Máster en Ciencia de la Computación en 1984.

"Profesor" por vocación, "historiador" por opción

En el balance retrospectivo aparece claramente mi vocación docente. Desde los primeros años en el CEC, en que tuve el privilegio de trabajar junto a un entrañable grupo de personas, sentí la necesidad de comunicar lo aprendido siguiendo el ejemplo de mis propios profesores en la tarea de contribuir a formar profesionales para esta nueva disciplina. Si bien el Departamento de Matemáticas tenía la tuición formal de la primera carrera de ingeniería en el área, en la práctica fue el CEC quien tomó el relevo del grupo de computación de matemáticas y asumió gradual e informalmente la responsabilidad por la docencia.

Con la creación del DCC, con el mismo director del CEC, la docencia para la IEPI tuvo continuidad y la carrera llegó a tener 400 alumnos. En ese contexto, mi llegada al DCC en 1979 fue la conclusión natural de mi vocación de "profesor" que continuó con la docencia en Plan Común, en IEPI, en el Bachiller en Computación y, a partir de los ochenta, en Ingeniería Civil en Computación. Y durante muchos años tuve la responsabilidad de la Coordinación Docente del DCC. Y en ese cargo, trabajamos en las reformas y renovaciones de los planes de estudios para sincronizarlos con los avances de la disciplina y con los estándares internacionales.

La preocupación y dedicación a la docencia, que no tenía el suficiente reconocimiento en la carrera académica, me llevaron a convertirlo en un tema de investigación. Las innovaciones docentes, especialmente en los cursos básicos de computación orientadas a centrar la docencia en el estudiante y su aprendizaje, dieron origen a publicaciones en congresos nacionales e internacionales en las áreas de Educación en Ingeniería y Educación en Computación. Por otra parte, y en los años

En el balance retrospectivo aparece claramente mi vocación docente. Desde los primeros años en el Centro de Computación [...] sentí la necesidad de comunicar lo aprendido siguiendo el ejemplo de mis propios profesores en la tarea de contribuir a formar profesionales para esta nueva disciplina.

recientes, contribuimos en la docencia de algunas de las nuevas universidades regionales públicas (de Talca, de O'Higgins y de Aysén) y en escuelas de verano para estudiantes y profesores de Educación Media.

Con el patrocinio de la Sociedad Chilena de Ciencia de la Computación (SCCC) que fue creada en 1984, en 1991 organizamos el "I Congreso Iberoamericano de Educación Superior en Computación" (CIESC) que prontamente fue acogido por el Comité que agrupa a las sociedades de computación latinoamericanas (CLEI) y hasta hoy se mantiene como uno de los eventos del congreso anual del CLEI (ver Figura 7). Posteriormente, y considerando el creciente interés por el tema, en 1998 creamos el "Congreso Chileno de Educación Superior en Computación" que es uno de los eventos de las jornadas anuales de la SCCC.

En el recuento de mis primeros años, no se puede dejar de mencionar el difícil contexto político y económico de la época. La dictadura intervino la Universidad y nombró rectores militares y redujo drásticamente el presupuesto con consecuencias en todos los ámbitos, tanto en la libertad de pensamiento y organización, como en los recursos para todas las actividades. Para defender la Facultad y la Universidad nos organizamos en la Asociación de Académicos, y, junto a académicos de otras universidades, en la Asociación Universitaria y Cultural Andrés Bello. Y en el ámbito profesional, desarrollamos proyectos informáticos de apoyo a la defensa de los derechos humanos y contribuimos a organizar la especialidad de Computación en el Colegio de Ingenieros de Ejecución en 1982 para defender, tanto a la carrera y a la profesión, como a las empresas y universidades del Estado.

En otro ámbito, y como egresado de una de las primeras generaciones de la primera carrera del área en Chile, y al comprobar que no había registro de la evolución de la disciplina en el país, sentí la obligación moral de investigarla y divulgarla. El trabajo ha dado origen a publicaciones nacionales e internacionales y a la organización de dos ediciones del "Taller de Historia de la Computación en Chile" (ver Figura 8) y a eventos conmemorativos de los principales hitos nacionales. Y, en asociación con investigadores latinoamericanos, hemos publicado y participado en los comités de organización y de programa del "Simposio de Historia de la Informática en América Latina y el Caribe" (SHIALC).

En síntesis, ingeniero de profesión, académico por ocupación, "profesor" por vocación e "historiador" por opción, interpretan mi involucramiento en el área de computación. Y en estas cuatro dimensiones formé parte de una red de colaboración con profesores, colegas, compañeros(as) de estudio y trabajo, incluyendo a las y los funcionarios del DCC y del CEC. El desarrollo de la disciplina fue y es un trabajo colectivo, de continuidades y cambios, y he tenido el privilegio de estar presente y contribuir en los saltos cuantitativos y cualitativos que explican el estado de la ciencia de la computación, y de su docencia en la universidad y en el país, y que ha permitido formar a generaciones de profesionales. ▪

Doctorados

Miguel Campusano

Título tesis: Mapping State Machines to Developers' Mental Model: Fast Understanding of Robotic Behaviors in the Real World

Profesores guías: Alexandre Bergel - Johan Fabry

Cuando hice mi pregrado nunca realicé ningún tipo de investigación ni se me pasó por la mente hacer un doctorado. Sin embargo, al tiempo de trabajar en un emprendimiento, me terminé aburriendo ya que los temas que abordábamos me parecían poco motivantes. Mi plan era volver donde hice mi pregrado, al Departamento de Ciencias de la Computación (DCC) de la Universidad de Chile, pero para hacer un magíster. Hablé con el profesor Johan Fabry, el mismo que me guió en mi pregrado, y, debido a diversos problemas que tenía en ese momento, me recomendó hacer un doctorado con un tema que era increíble para mí en ese entonces: robótica. Debo admitir que mi mente se iluminó de inmediato, en el DCC nunca tuve contactos con robots.

Mi paso por el doctorado fue, por decirlo de alguna forma, complejo. Al comienzo todo era emocionante, eso es lo que pasa cuando uno aprende algo totalmente nuevo y fascinante (recordemos que nunca tuve un paso por investigación antes de eso). Luego, el camino se volvió bastante tortuoso, no fue fácil para mi encarar el mundo de la academia e investigación, y mi salud mental se vio afectada bastante (al parecer un tema más que conocido en este mundo y, por alguna razón, tabú). Sin embargo, aprendí a golpes a cómo llevar este proceso y, finalmente, supe llevar mi proyecto de investigación. Además, justo en medio de mi doctorado Johan tuvo que dejar la Universidad y ahí Alexandre Bergel me tomó bajo su tutela. No fue un proceso fácil, pero le agradezco enormemente a Johan y Alex el ayudarme en esta carrera y darme la confianza que necesitaba para llevar el doctorado. Claramente, como estudiantes, nos faltan grupos de ayuda para que podamos llevar esta carrera de forma saludable.

Mi tema de doctorado consistió en desarrollar un lenguaje de programación para comportamientos robóticos con una característica en particular, el robot se mueve al mismo tiempo que se está programando. A esto se le conoce como programación en vivo. El objetivo de este lenguaje es hacer más fácil el desarrollo de comportamientos robóticos. Este tipo de unión entre una disciplina netamente ligada a computación (programación en vivo) con la robótica fue algo novedoso en su tiempo y que, afortunadamente, he visto como va aumentando en popularidad, con workshops y conferencias dedicadas sólo a unir la computación y la robótica. Me alegra ver que un tema tan importante como éste, que muchas veces es dejado de lado, esté siendo tomado en cuenta y mucha gente le esté dedicando el tiempo que merece.

Aunque mi trabajo con este lenguaje de programación me enseñó mucho sobre el proceso de programar robots, al evaluarlo no pudimos comprobar nuestra hipótesis, no podemos afirmar que el lenguaje facilita, de alguna forma, el desarrollo de comportamientos robóticos. Aun así creo que vamos en la dirección correcta, programar un robot requiere la integración de diferentes disciplinas, todas sumamente complejas. No sólo se van a producir robots más complejos a través de mejorar la inteligencia artificial, algoritmos de control, visión computacional, etc., sino también es importante ayudar a que los programas robóticos sean más fáciles de escribir y de integrar al robot mismo, y con más capacidades. Todo esto para hacer comportamientos robóticos cada vez más complejos y útiles para la sociedad.

El tema de la robótica me lleva hoy en día a investigar y diseñar arquitecturas para drones, haciendo un postdoctorado en la Universidad del Sur de Dinamarca (SDU), en el marco del proyecto HealthDrone. En este proyecto queremos transportar medicinas y otros artículos médicos entre diferentes hospitales y centros médicos que pueden estar ubicados en zonas de difícil acceso, incluyendo islas donde sólo se puede llegar en barco. La idea es usar drones para reducir el costo y el tiempo de traslado de estos artículos médicos.

Aunque estoy en Dinamarca no me he desligado del mundo de la robótica en Chile. Hemos iniciado (con otras personas ligadas a la robótica) una corporación sin fines de lucro llamada Cuac. Con esta corporación trabajamos para potenciar la robótica y su educación en Chile.

Matías Toro

Título tesis: Abstracting Gradual Typing: Metatheory and Applications

Profesor guía: Éric Tanter

Egresé de ingeniero civil en computación de la Universidad de Chile el año 2007, y luego me dediqué a trabajar en la industria aeronáutica por siete años. El trabajo, a pesar de no ser trivial, resultaba monótono y a veces tedioso. No me imaginaba todo el resto de mi vida haciendo lo mismo. Buscando nuevos desafíos, y dado que por temas familiares me complica salir al extranjero, el año 2013 volví al Departamento de Ciencias de la Computación (DCC) para realizar el Magíster en Ciencias, mención Computación. Es ahí donde conocí a mi profesor guía Éric Tanter, el cual me reintrodujo al área de lenguajes de programación. Me gradué del magíster en el 2013, y en el 2014 se hizo natural extender mi trabajo hacia un doctorado.

Mi tesis de doctorado se llama "Abstracting Gradual Typing: Metatheory and Applications", y la investigación se centró en los lenguajes de programación graduales, los cuales buscan integrar sistemas de tipos estáticos (como el de Java) con sistemas de tipos dinámicos (como el de Python). Con sistemas de tipos graduales el programador puede escoger qué expresiones anotar con información de tipos estática, y cuáles dejar sin especificar. El sistema de tipos gradual chequea en tiempo de ejecución lo que no puede verificar durante la etapa de compilación, asegurando así que no se violen las anotaciones estáticas.

El enfoque clásico para diseñar lenguajes graduales es usualmente *ad-hoc*, pero existen metodologías que sistematizan este proceso. Una de ellas es *Abstracting Gradual Typing* (AGT), que ayuda a construir sistemáticamente lenguajes graduales a partir de lenguajes estáticamente tipados usando interpretación abstracta. Mi trabajo de investigación exploró esta (casi nueva en ese entonces) metodología, aplicando AGT a distintas disciplinas de tipo y mecanismos de lenguajes complejos.

La mayor parte de mi investigación fue teórica y se puede resumir en lo siguiente. Se partía de un sistema de tipos complejo existente, que satisfacía cierta propiedad formal, se aplicaba sistemáticamente AGT, y luego se observaba si el lenguaje gradual resultante también cumplía o no con dicha propiedad. Éste no fue el caso de todos los lenguajes estudiados, por lo que se tenía que ir modificando ciertas abstracciones y reglas de evaluación, para que pudiera cumplir con la propiedad, sin perder otras propiedades intrínsecas a los lenguajes graduales. Estas iteraciones conllevaron muchas demostraciones matemáticas distintas, las que consumieron la mayor parte del tiempo de mi doctorado.

Toda esta experiencia fue una montaña rusa de emociones, donde uno se esperanzaba de tener una nueva idea o solución que luego se derrumbaba al encontrar algún problema en alguna demostración de algún lema. Trabajar por meses en una demostración matemática para luego ver que había un error (a veces a días antes del *deadline* de una conferencia), sumada a la presión de terminar a tiempo el doctorado fue muy estresante. Llegué a soñar con demostraciones (y hasta encontré algunos errores en demostraciones así).

La otra parte difícil fue la de escribir *papers*, ya que para hacer investigación no sólo sirve ser bueno técnicamente, sino que también se debe saber transmitir las ideas. Aprendí que escribir un *paper* puede ser muy parecido a desarrollar un software: no es recomendable partir ciegamente, sino que hay que darle estructura a las ideas y planificar cómo se van a presentar las cosas de manera de que todo fluya. También aprendí a usar otro tipo de inglés empleado en artículos científicos, que es distinto al que uno podría estar acostumbrado a leer.

Cuando hice el doctorado no había un curso que te enseñara todo esto y lo que aprendí, lo aprendí de Éric. Sigo sintiendo que es mi punto débil y que me falta mucho por aprender aún. Relacionado con esto, también tuve que aprender a presentar artículos científicos. Muchas veces gastaba semanas preparando e iterando una presentación. A pesar de todas las dificultades, es muy gratificante finalmente llegar a publicar un *paper* y presentarlo. ¡Vale la pena el esfuerzo!

Actualmente me encuentro haciendo un postdoctorado en el DCC continuando mis temas de investigación pero en temas relacionados con privacidad diferencial en lenguajes de programación.

Mauricio Quezada

Título tesis: Knowledge Discovery from News Events on Twitter

Profesora guía: Bárbara Poblete

Estudié Ingeniería Civil en Computación en el Departamento de Ciencias de la Computación (DCC) de la Universidad de Chile. Mi gusto por los distintos temas que vi en los cursos que tuve durante la ingeniería, más la buena relación que he tenido con algunos profesores del Departamento me llevaron a continuar con un magíster, y luego, con el doctorado.

Mi tesis de doctorado consistió en una exploración de distintas formas de extraer conocimiento desde la información que comparten los usuarios de Twitter sobre eventos noticiosos. Estas formas de extracción se basan en la suposición de que el contexto en que se publica esta información es muy importante para agrupar contenido similar. Por ejemplo, uno de los trabajos consistió en representar los tuits que expresan algún comentario relacionado a un evento noticioso en particular, como la muerte de Nelson Mandela en 2013, como la diferencia de tiempo en que fueron publicados dos mensajes consecutivos. Esta simple representación nos permitió observar que cierto tipo de noticias generan mayor actividad de los usuarios, y que los mensajes que publican en este tipo de noticias son muy distintos a los mensajes que publican sobre noticias con menores niveles de actividad. Otro aspecto importante es que un mensaje individual no dice mucho sobre la noticia, pero el considerar una gran cantidad de ellos nos permite observar patrones interesantes. Este trabajo lo realicé durante mi tesis de magíster —y luego profundizado durante el doctorado—, en conjunto con Janani Kalyanam y Gert Lanckriet, en ese entonces de la Universidad de California, San Diego.

Una de las cosas más complicadas que enfrentamos durante el desarrollo de mi tesis fue la falta de conjuntos de datos "correctos" sobre los cuales hubiéramos podido evaluar nuestros modelos. Debido a la gran variabilidad de la información (una misma noticia no ocurre dos veces) y la naturaleza de los modelos que propusimos (orientados a resolver nuevas tareas en la minería de datos) nos exigió pensar en formas novedosas y válidas de evaluar la metodología. En pocas palabras, la evaluación consistió en identificar que nuestros modelos hacían resaltar patrones interesantes en otros aspectos de los datos.

Lo más desafiante del doctorado fue poder gestionar mi tiempo y definir bien los objetivos de cada etapa. Siempre aparecían nuevas ideas o cosas interesantes en las que trabajar, por lo que definir bien el plan —y uno no sabe qué va a encontrar al final— fue complicado. Por otro lado, creo que lo más interesante ha sido poder desarrollar distintas habilidades con el tiempo; simplemente el tener la experiencia de trabajar en investigación va generando nuevas capacidades que uno empieza a notar hacia el final del doctorado. También el poder hacer clases en distintas instancias fue muy gratificante, aunque estresante, ya que tuve la oportunidad de transmitir lo que he ido aprendiendo.

Decidí no seguir una carrera académica por varios motivos. Actualmente soy cofundador y CTO de Cero.ai, una empresa que automatiza procesos de comunicación entre empresas y personas.

Daniel Hernández

Título tesis: The Problem of Incomplete Data in SPARQL

Profesor guía: Claudio Gutiérrez

Cuando estaba el colegio, aún sin decidir que estudiaría, mi interés era estudiar algo que me permitiera poder entender el mundo, y por ello pensaba que cualquier carrera que tuviera que ver con ciencias me podría gustar. Escogí entrar a la Escuela de Ingeniería de la Universidad de Chile porque tenía un Plan Común que conducía a muchas carreras, lo que me permitiría más tarde decidir qué estudiar. Al final me decanté por computación, pues me gustaba y a la vez sentía que era una ciencia bastante general. Luego de hacer un magíster y un año de trabajar haciendo clases en la Universidad de Talca, volví al Departamento de Ciencias de la Computación (DCC) donde había estudiado, esta vez para hacer un doctorado. Me decidí a ello porque disfruté el año que trabajé haciendo clases y también por la recomendación de Claudio Gutiérrez (mi profesor guía).

Lo más complejo de mi doctorado fue sin duda el proceso de maduración que se produce cuando uno pasa de intentar resolver un problema a entender cuál es el problema que uno está resolviendo, y el impacto que puede tener lo que uno está haciendo. Este proceso va acompañado con lo difícil que resulta comunicar los resultados de la investigación, escribirlo de manera clara y siguiendo las prácticas de otros investigadores. Como he terminado mi doctorado hace poco tiempo, puedo recordar el camino que he seguido y percibir el cambio que se produce en este proceso de maduración. Lo que hoy me resulta evidente, antes no lo era.

Lo otro que requiere esfuerzo es mantenerse focalizado. Para investigar uno tiene que simplificar al máximo el problema abordado. Quitarle todos sus aspectos no esenciales hasta que el problema sea lo suficientemente claro como para poder enfrentarlo y luego poder comunicarlo. Hacer esto no es fácil. Al simplificar un problema uno termina generando una larga lista de variantes y preguntas sin resolver, para retomar algún día. También resulta un poco desalentador estar enfocado en un problema que se hace cada vez más pequeño al lado de la larga lista que voy dejando al lado, va un poco contra la motivación inicial de comprender el mundo. Esto produce la sensación de que uno cada vez sabe menos. Por suerte, a lo largo del doctorado me hice consciente de este fenómeno, lo que ahora me ayuda a lidiar con ello.

Mi relación con Claudio fue siempre muy buena y puedo decir que aprender de su mirada general fue una de las cosas más positivas del doctorado. También tengo que agradecer a Aidan Hogan y Renzo Angles, de quienes también aprendí sus diferentes miradas cuando trabajábamos en algún *paper*. El ambiente del DCC es muy bueno para hacer un doctorado, porque tiene una comunidad amable con la cual compartir y reflexionar, y gente muy admirable.

Mi tesis de doctorado "The Problem of Incomplete Data in SPARQL", estudia cómo las nociones de la información incompleta se manifiestan en el lenguaje de consulta SPARQL. Este lenguaje fue definido por el World Wide Web Consortium (W3C) para los datos de la Web, en particular, para lo que se conoce como Web Semántica. El modelo de datos de SPARQL, llamado RDF, fue diseñado teniendo en consideración que la Web es un espacio en el cual múltiples actores publican de manera independiente, con diferentes creencias y maneras de modelar (o entender) el mundo. Esto nos lleva a que todos los conjuntos de datos en la Web sean considerados incompletos. Por el contrario, SPARQL es un lenguaje que surge (varios años después de RDF) de la necesidad de explorar un conjunto acotado de datos RDF, es decir, de la manera tradicional. Esta diferencia entre RDF y SPARQL produce incompatibilidades entre ambos lenguajes.

Para entender mi trabajo creo que es necesario revisar la noción filosófica que tenemos de la noción de "entender". Desde el punto de vista científico creo que el concepto de "entender" está relacionado con aquello que ocurre cuando uno analiza un conocimiento bajo una formulación o teoría diferente de la original. En mi tesis, yo tomo las definiciones del lenguaje SPARQL y las analizo bajo la teoría de información

incompleta para bases de datos. La formulación de SPARQL consiste en una serie de reglas definidas de forma semiformal que describen una función que toma una base de datos en lenguaje RDF y una consulta en lenguaje SPARQL y entrega un conjunto de soluciones. Por otra parte, la semántica del lenguaje RDF consiste en asociar cada base de datos con un conjunto de posibles modelos del mundo representado. Bajo la teoría de información incompleta de las bases de datos, la pregunta natural es si la semántica de SPARQL es consistente con la semántica de RDF. Una definición concreta de esto es, por ejemplo, saber si las soluciones que se entregan para una consulta y una base de datos dadas son aún válidas para todos los modelos del mundo que la base de datos representa. A las soluciones que poseen tales características se las conoce como *certain answers* o soluciones seguras.

Para analizar el problema de las soluciones seguras en SPARQL tomé en consideración un fragmento de SPARQL con una semántica bien definida y una simplificación de la semántica de RDF que considera a los datos como sentencias con variables (y por ende incompletas). Por ejemplo, una sentencia como "Juan tomó el bus desde Santiago a *x*" es incompleta porque, si bien sabemos que el bus que Juan tomó tenía un lugar de destino, no sabemos cuál era. El lenguaje RDF tiene un elemento que coincide exactamente con esta noción de variable: los "nodos blancos". Usando esta simplificación podemos formular la pregunta: ¿Produce SPARQL soluciones que no sean seguras? La respuesta es afirmativa. Una solución que no es segura se produce, por ejemplo, si la base de datos dice que "Juan tomó el bus de Santiago a *x*" y la respuesta a la consulta "¿a qué lugar Juan no tomó el bus?" incluye a Curicó como respuesta. Esta respuesta es insegura porque en un mundo posible la variable *x* puede tomar el valor Curicó.

La pregunta que sigue es cómo podemos modificar la semántica de SPARQL para obtener sólo respuestas seguras. Una semántica de dichas características debe considerar que el problema de si una solución es segura está en la clase de complejidad coNP (muy complejo), mientras que el fragmento SPARQL de nuestra formulación se puede computar de una forma muy eficiente (AC°). Entonces, una parte de mi tesis consistió en proponer y evaluar experimentalmente la factibilidad práctica de un método aproximado para la evaluación de SPARQL, que entrega sólo respuestas seguras, pero que algunas veces no las entrega todas.

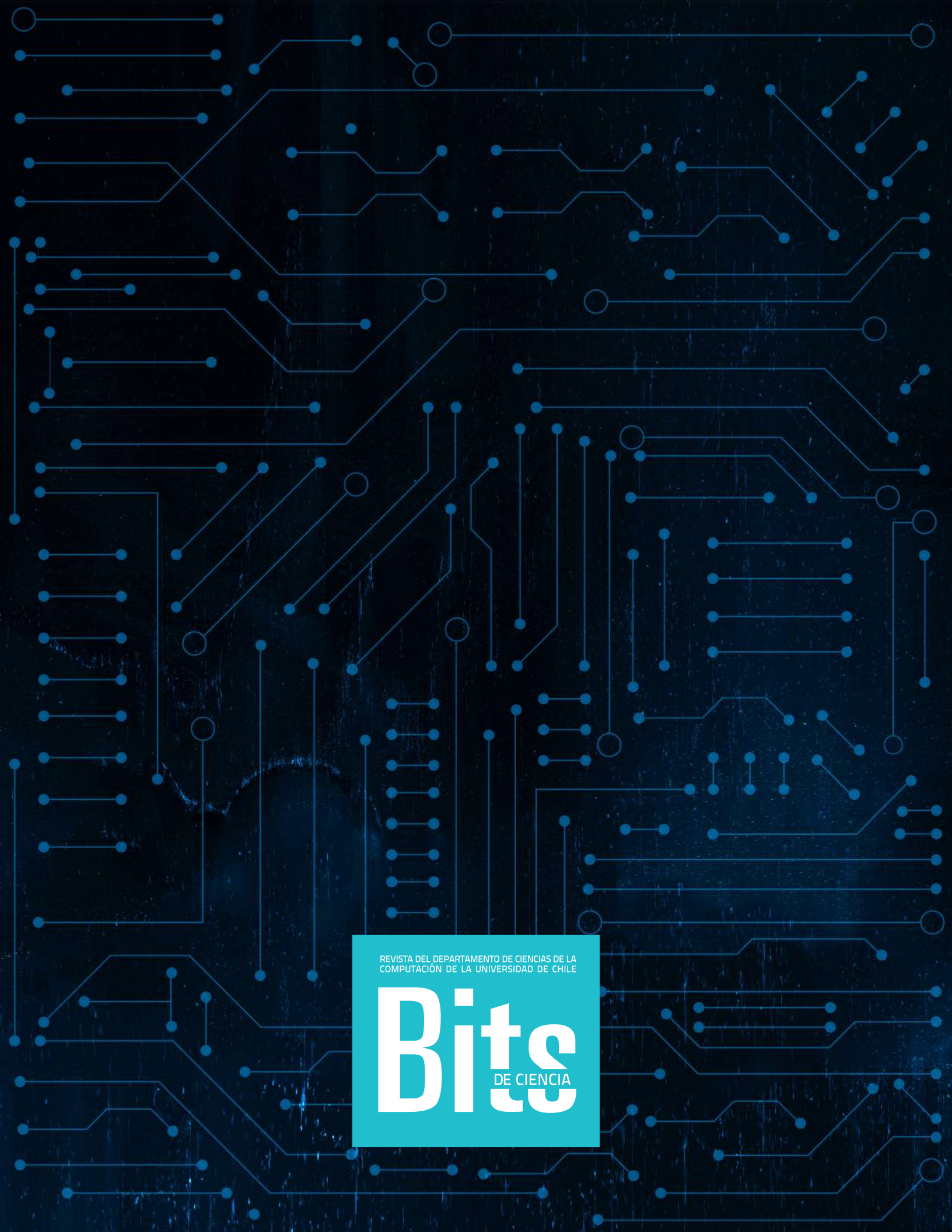
Actualmente, estoy trabajando en la Universidad de Aalborg, en Dinamarca, como postdoc en DAISY - Center of Data Intensive Systems.

dcc
CIENCIAS DE LA COMPUTACIÓN
UNIVERSIDAD DE CHILE

Nano Break

Podcast del Departamento de Ciencias de la Computación de la Universidad de Chile

Disponible en: DCC UChile Nano Break



REVISTA DEL DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN DE LA UNIVERSIDAD DE CHILE

Bits DE CIENCIA