

EDITORIAL

Data Science: towards automated decision making

Analyzing data today is a highly relevant task, as data is an asset of great value for both researchers and organizations. As a result of the advancement in information and communication technologies, today there are many platforms that allow data to be recorded. For example, the information systems that support organizations in managing their processes record events, such as purchases, money transfers, medical visits, product delivery, notarial procedures, etc., in databases. Every day, electronic commerce moves millions of dollars in transactions and all the details of each transaction are recorded in the process. Social media generates petabytes of data daily. Facebook alone has more than 1 billion active users who contribute hundreds of petabytes of data.

At the beginning of 2000s, researcher Doug Laney proposed that what makes analyzing such vast amounts of data challenging is what he described as the three “V’s” of Big Data: Volume, Velocity and Variety. Already in the same decade, it was projected that the growing incorporation of new information technologies would result in an exponential increase in the amount of data requiring analysis. Analyzing the data that is generated with these technologies allows organizations to have greater knowledge allowing better decision making. However, extracting knowledge from data is not an easy task, since it is a process that depends on the “V’s” mentioned above. Processing data is crucial, as it is what turns information, values and past experiences into knowledge. Applying knowledge, in turn, generates new data that requires processing, creating a cycle that repeats itself infinitely.

Nowadays, the biggest obstacle that organizations face in this cycle is not how to analyze data, but rather how to extract knowledge from it and apply it in order to improve business¹ (LaValle, Lesser, Shockley, Hopkins and N. Kruschwitz, 2011). This is why, in the last two decades, new research areas have emerged within Data Science, which contribute with data processing and analysis tools and techniques. Data Science is now an interdisciplinary area whose purpose is to transform data into value, in order to report, diagnose, predict and also recommend solutions or improvements in products, services and/or processes².

In the decades following their emergence as disciplines, the impact of Artificial Intelligence and Machine Learning on advancement and knowledge was fluctuating. Today, however, thanks to greater maturity of technology and to developments in Data Science, it is possible to advance more quickly thanks to certain factors, for example: improvements in computer architecture that allow greater storage and computing capacity; complex, open source libraries that allow researchers to focus on data processing analysis rather than infrastructure; the availability of open data³ that is the main input of mathematical models to better understand reality.

There is no doubt that the past decade has been characterized by advances in Artificial Intelligence. Among the many projects that have had the most relevance in the community are: Zippedi, a Chilean robot that scans supermarket aisles to replenish stock and is now in the process of creating a platform

¹ S. LaValle, E. Lesser, R. Shockley, M.S. Hopkins and N. Kruschwitz. “Big data, analytics and the path from insights to value”. MIT sloan management review. Vol. 52, Issue 2, pp. 21-32. 2011.

² W. Van Der Aalst. “Data science in action. In Process mining”. Springer, Berlin, Heidelberg, pp. 3-23. 2016.

³ <https://datos.bcn.cl/es/informacion/mapa-global-de-datos-abiertos>

that allows online purchases; navigation systems that use GPS to optimize routes; voice assistants like Google Home or Amazon Echo, etc.

It is vital for organizations to be able to provide a timely response to their clients' requirements, whether related to products or services. Through the implementation of data analysis models, it is possible to reduce the cognitive effort involved in retrieving and processing information to later make decisions. The effort must be put into managing information in an automated way and generating models that allow automated decisions to be made that deliver accurate and timely responses. In this sense, there are two things research should be oriented towards. First, the extraction of knowledge from unstructured data (an area of study known as Natural Language Processing), which allows taking customer requirements, processing them automatically and giving immediate response. Second, image processing using neural networks as a computational model that makes it possible to find patterns with a probability percentage allowing great advances in areas such as computer vision and voice recognition⁴.

The challenges that arise in the automated decision making are knowing the usability limit of the data and being able to answer questions such as: Who is the owner of the data? Can the integrity of the data be guaranteed? What is the limit on the use of data by organizations? Which data is private? There are more questions that could be raised, however, the main discussion is how to generate better automated services through Machine Learning and descriptive and predictive data analysis where decisions have a low cost and risk.

Mauricio Arriagada-Benítez, PhD.
Facultad de Ingeniería,
Departamento de Ingeniería en
Computación e Informática,
Universidad de Tarapacá.
Arica, Chile.

E-mail: mauricio.arriagada@academicos.uta.cl

⁴ J.J. López Murphy y G. Zarza. "La Ingeniería del Big Data: Como Trabajar Con Datos". Print. Tecnología. Vol. 534. 2017.

EDITORIAL

Ciencia de Datos: hacia la automatización de las decisiones

Analizar datos hoy en día es una tarea sumamente relevante por cuanto es un activo de gran valor tanto para investigadores como para organizaciones. Gracias al avance en tecnologías de la información y comunicación, hoy en día existen muchas plataformas que permiten registrar datos. Por ejemplo, los sistemas de información que apoyan a las organizaciones en la gestión de sus procesos registran todos los eventos en bases de datos tales como compras, traspasos de dinero, visitas médicas, entrega de productos, trámites notariales, etc. El comercio electrónico mueve millones de dólares diariamente en transacciones y con ello todos el detalle de cada transacción. Las redes sociales generan petabytes de datos diarios. Solo la red social Facebook tiene más de 1 billón de usuarios activos quienes aportan con cientos de petabytes de datos.

A comienzos del 2000, el autor Doug Laney introduce el términos de las ‘V’ que caracterizan a Big Data y que son los desafíos a estudiar: Volumen, Velocidad y Variedad de los datos. Ya en la primera década del 2000 se proyectaba el aumento en datos que produciría como resultado de la incorporación de nuevas tecnologías de información. Al analizar los datos que se generan con estas tecnologías, permite a las organizaciones tener mayor conocimiento permitiendo una mejor tomar de decisiones. Extraer conocimiento de los datos no es una tarea sencilla, ya que requiere de un proceso que depende las ‘V’ antes mencionadas. El procesar los datos genera información con un valor agregado que es el insumo para el descubrimiento de conocimiento que se basa en la información, valores y también en experiencia y cómo se hacen la cosas. Esto es un ciclo que se repite ya información y conocimiento a su vez generan más datos.

Hoy en día el mayor obstáculo que presentan las organizaciones no son los datos sino la comprensión de éstos para aplicar, análisis y mejorar los negocios¹ (LaValle, Lesser, Shockley, Hopkins y Kruschwitz, 2011), es por ello que en las últimas décadas han surgido áreas de investigación aportando con herramientas y técnicas de procesamiento y análisis de datos. Es así como Ciencia de Datos es una área interdisciplinaria que tiene como propósito transformar datos en valor² para poder reportar, diagnosticar, predecir y también recomendar soluciones o mejoras en productos, servicios y/o procesos.

Enmarcado dentro de Ciencia de Datos, Inteligencia Artificial y Machine Learning, desde que emergen como disciplina, han tenido un fluctuante impacto en desarrollo y conocimiento. Sin embargo, hoy en día, gracias a mayor madurez en tecnología y avances en la Ciencia de Datos, es posible avanzar con mayor firmeza mediante el cumplimiento de ciertas condiciones. Por ejemplo, mejoras en arquitectura de computadores que permiten mayor capacidad de almacenamiento y computo, librerías de código abierto de gran complejidad que permiten a los investigadores centrarse en análisis de procesar datos más que en la infraestructura y algo muy importante, la disponibilidad de datos abiertos³ que son la principal entrada de modelos matemáticos para comprender mejor la realidad.

No hay duda alguna que la pasada década se caracterizó por avances en Inteligencia Artificial. Entre los muchos proyectos que más relevancia han tenido en la comunidad destacan Zippedi, robot chileno

¹ S. LaValle, E. Lesser, R. Shockley, M.S. Hopkins and N. Kruschwitz. “Big data, analytics and the path from insights to value”. MIT sloan management review. Vol. 52, Issue 2, pp. 21-32. 2011.

² W. Van Der Aalst. “Data science in action. In Process mining”. Springer, Berlin, Heidelberg, pp. 3-23. 2016.

³ <https://datos.bcn.cl/es/informacion/mapa-global-de-datos-abiertos>

que permite escanear pasillos de supermercados para reposición stock y ahora en proceso creación de una plataforma que permite realizar compras online, sistemas de navegación que usan GPS en optimización de rutas, asistentes de voz como Google Home o Amazon Echo, entre muchos otros.

Es de vital importancia que las organizaciones puedan dar respuesta oportuna a los requerimientos de sus clientes, ya sea en productos o servicios. A través de la implantación de modelos de análisis de datos, es posible reducir el esfuerzo cognitivo que implica recuperar y procesar información para luego tomar decisiones. El esfuerzo debe estar puesto en gestionar información de manera automatizada y generar modelos que permitan tomar decisiones automatizadas que entregan respuestas precisas y oportunas. En este sentido las líneas de investigaciones deben estar orientadas en la extracción de conocimiento de datos no estructurados (área de estudio conocida como Procesamiento del Lenguaje Natural) que permita tomar requerimientos de clientes, procesarlos automáticamente y dar respuesta inmediata, el procesamiento de imágenes que busca patrones usando redes neuronales como modelo computacional que permite encontrar patrones con un porcentaje de probabilidad logrando grandes avances en áreas como visión por computador y reconocimiento de voz⁴.

Los desafíos que se presentan en decisiones automatizadas es saber el límite de usabilidad de los datos y poder responder preguntas como ¿quién es el dueño de los datos?, ¿se puede garantizar la integridad de los datos?, ¿cuál es límite en el uso de datos por parte de las organizaciones?, ¿qué datos son privados?. Son varias las preguntas que se podrían seguir planteando, sin embargo, la discusión central es cómo generar mejores servicios automatizados a través de aprendizaje automático y un análisis de datos tanto descriptivo, como predictivo donde las decisiones tengan un bajo costo y riesgo.

Mauricio Arriagada-Benítez, PhD.

Facultad de Ingeniería,

Departamento de Ingeniería en

Computación e Informática,

Universidad de Tarapacá.

Arica, Chile.

E-mail: mauricio.arriagada@academicos.uta.cl

⁴ J.J. López Murphy y G. Zarza. "La Ingeniería del Big Data: Como Trabajar Con Datos". Print. Tecnología. Vol. 534. 2017.