

Analíticas para Sistemas de Atención con Grandes Volúmenes de Eventos

Esteban Schab^(1,3), Carlos Casanova^(1,3) y Fabiana Piccoli^(1 y 2)

⁽¹⁾ Universidad Autónoma de Entre Ríos, Concepción del Uruguay

⁽²⁾ LIDIC- Univ. Nacional de San Luis, San Luis

⁽³⁾ Univ. Tecnológica Nacional, Facultad Regional Concepción del Uruguay, Concepción del Uruguay
Argentina

mpiccoli@unsl.edu.ar

{schabe, casanovac}@frcu.utn.edu.ar

RESUMEN

Cualquier proceso de negocio, hoy en día, exige tomar decisiones rápidas, las cuales se deben ir adaptando a los cambios. Una posibilidad es resolver el proceso computacionalmente. Para ello es necesario, en base a los datos, producir analíticas, las cuales buscan transformarlos en conocimiento para la toma de decisiones. Existen varios tipos de analíticas, particularmente en este trabajo proponemos una línea de investigación enfocada en la analítica prescriptiva, como nivel más avanzado, capaz de calcular acciones a ser ejecutadas en el momento (decisiones operativas) o en el futuro (decisiones tácticas para corto y mediano plazo, decisiones estratégicas para largo plazo) para lograr un objetivo deseado. El cálculo de las acciones involucra otros aspectos, uno de ellos derivado de la necesidad de bajos tiempos de respuesta, la computación de alto desempeño, particularmente involucrada en el procesamiento de *datastreams*.

Palabras clave: Inteligencia Computacional. Analíticas. Big Data. Computación de Alto Desempeño.

CONTEXTO

Esta propuesta de trabajo se lleva a cabo dentro de los proyectos de investigación: “Tecnologías Avanzadas aplicadas al Procesamiento de Datos Masivos” (LIDIC, UNSL), “Cómputo de Altas Prestaciones aplicado a la Solución de Grandes Problemas” (UADER) y “Descubrimiento de conocimiento en bases de datos” (GIBD, UTN).

1. INTRODUCCIÓN

La mejora continua y adaptativa de los procesos de negocio resulta clave para mantener la competitividad de las organizaciones. La digitalización de los procesos, así como el incremento en las tecnologías de monitoreo, han llevado a producir una enorme cantidad de datos, los cuales tienen un gran potencial para la mejora de los procesos conducida por analíticas [3] [5] [15].

Las analíticas buscan transformar los datos en conocimiento para la toma de decisiones [4], pudiendo distinguirse cuatro tipos de analíticas según el nivel de automatización del proceso [7]. Primero, la descriptiva intenta responder qué ha pasado o está pasando y la diagnóstica

por qué ha pasado o está pasando. Ambas analizan datos históricos. Luego, la analítica predictiva busca responder qué sucederá. Aplica el conocimiento para predecir nuevos datos sobre el presente o el futuro (pronóstico). Ninguno de estos enfoques sugiere acciones concretas, sino que descansan en el juicio subjetivo y las habilidades analíticas del usuario para deducir acciones de mejora. Finalmente, la analítica prescriptiva responde qué debería hacerse para lograr un objetivo. Determina acciones a ser ejecutadas en el momento (decisiones operativas) o en el futuro (decisiones tácticas para corto y mediano plazo, decisiones estratégicas para largo plazo).

A pesar de los avances tecnológicos en general, las analíticas de procesos existentes dentro de la industria actual no aprovechan completamente el conocimiento oculto en los grandes volúmenes de datos con los que cuentan debido a las siguientes limitaciones[3]:

- a. No hacen uso de técnicas prescriptivas para transformar los resultados del análisis en acciones de mejora concretas, dejando este paso completamente a criterio del usuario.
- b. Hacen un uso intensivo de datos de sistemas en producción, generando un deterioro en el desempeño de las herramientas de software que soportan los procesos.
- c. La optimización es conducida *ex post*, después de completado el proceso, en contraste a la mejora proactiva durante la ejecución del proceso.

Dentro de *Big Data* puede, también, identificarse como área emergente el procesamiento de *datastreams*, o llamado *Data Stream Mining* [11] [17]. Un *datastream* es una representación digital y transmisión continua de datos que describen una clase de eventos relacionada [13]. Su procesamiento permite

lograr respuestas en tiempo real a los eventos en forma de toma de decisiones.

En los sistemas de atención, particularmente los *datastreams* abren nuevas y amplias oportunidades para la creación de valor en las organizaciones. Un caso de particular interés son los bancos o cualquier otros sistemas de atención al público [9]. En ellos la generación de *datastreams* es causada por los sistemas de gestión de la atención que implementan. Estos sistemas en general utilizan modelos relacionales de bases de datos y, si bien permiten la elaboración de analíticas, puede resultar inapropiada su implementación debido a que el desempeño del sistema de atención, especialmente en un contexto de *Software as a Service* (SaaS)[16], puede verse colapsado por las continuas consultas necesarias para realizar el monitoreo. Por tal motivo, la generación y el procesamiento de los eventos en forma de *datastreams* como una componente paralela al sistema de atención puede ser la tecnología de base para permitir un monitoreo eficiente sin el deterioro de la performance del propio sistema. En la siguiente sección se describen las características básicas de la línea de investigación a seguir.

2. LÍNEAS DE INVESTIGACIÓN, DESARROLLO E INNOVACIÓN

La línea de investigación se basa en la idea de la Optimización de procesos de negocio por recomendación.

Los grandes volúmenes de datos generados por los sistemas de atención pueden ser utilizados en algoritmos de aprendizaje supervisado o no supervisado para la generación de analíticas, particularmente las descriptivas y predictivas. Existe, sin embargo, una dificultad en lo relativo a las analíticas

prescriptivas: no cuentan con un “profesor” que les diga qué acción tomar en cada circunstancia. Un tipo de aprendizaje que no necesita de un profesor es el llamado aprendizaje por refuerzo. En este esquema es el propio agente quien es capaz de juzgar y criticar sus acciones con base en sus percepciones y de alguna medida de aptitud, recompensa o refuerzo. La tarea del aprendizaje por refuerzo es usar recompensas observadas para aprender una política óptima (o aproximadamente óptima) del entorno, sin asumir ningún conocimiento *a priori* [13]. Esta política le dice al agente qué hacer en cada estado posible a alcanzar. Por lo tanto, se juzga al aprendizaje por refuerzo como una herramienta muy útil para la elaboración de analíticas prescriptivas, ya que ambos comparten el mismo objetivo.

Esta línea propone, en consecuencia, la composición de modelos de analítica prescriptiva que superen los inconvenientes descritos. Estos modelos formarían una parte esencial de un proceso de mejora continua basado en la recomendación de acciones operativas y tácticas destinadas a mantener los indicadores de rendimiento del sistema dentro de los valores deseados.

Para la construcción de modelos prescriptivos, cuya principal función resulta en la determinación de las acciones a llevar a cabo, se debe hacer uso de modelo predictivos para explorar los futuros cercanos, modelos descriptivos para calcular la aptitud de dichos estados, y un lenguaje de especificación que permita determinar valores deseados y estrictos de aptitud, es decir, una medida de calidad de sistema. En consecuencia, para complementar el aprendizaje por refuerzo se apela a técnicas provenientes de la inteligencia computacional: redes neuronales como modelos, teoría de conjuntos difusos como lenguaje de especificación, y métodos numéricos y

metaheurísticos para entrenamiento de tales modelos [2] [14] [17].

Como las analíticas se encuentran inmersas dentro de un proceso de negocio dinámico, que exige tiempos de respuestas rápidos, y dadas las características propias de cada una de las técnicas de inteligencia computacional antes mencionadas, es mandatorio pensar en la aplicación de modelos/paradigmas de computación de alto desempeño[1][6][10], principalmente en el entrenamiento y ejecución de analíticas. Este proceso debe ser lo suficientemente rápido como para procesar los *datastreams* que el sistema genera de manera continua y brindar resultados en tiempo real.

3. RESULTADOS OBTENIDOS/ESPERADOS

En cuanto a las analíticas, se espera construir un modelo prescriptivo dirigido por los datos que constituya una parte esencial de un proceso de mejora continua. Este modelo estará basado en la recomendación automática y proactiva de acciones operativas y tácticas destinadas a mantener los indicadores de rendimiento de un sistema de atención dentro de los valores deseados, y en un contexto con grandes volúmenes de eventos.

En cuanto al procesamiento, se espera construir y probar algoritmos de procesamiento aplicando computación paralela capaz de reducir los tiempos derivados de entrenar y ejecutar el modelo. Para ello es deseable facilitar el desarrollo de una solución paralela portable, de costo predecible, capaz de explotar las ventajas de modernos ambientes HPC a través de herramientas y “frameworks de computación” de alto nivel.

4. FORMACIÓN DE RECURSOS HUMANOS

Los resultados esperados respecto a la formación de recursos humanos son el desarrollo de 1 tesis de doctorado, 2 de maestría y de varias tesinas de grado en las universidades intervinientes.

5. BIBLIOGRAFÍA

- [1] Mercedes Barrionuevo, Mariela Lopresti, Natalia Miranda, and María Fabiana Piccoli. “Solving a big-data problem with gpu: the network traffic analysis”. *Journal of Computer Science and Technology*, 15(01): p.30–39, Apr. 2015.
- [2] A. Ebrahimnejad and J. L. Verdegay. “Fuzzy sets-based methods and techniques for modern analytics”. Springer International Publishing. 2018.
- [3] Christoph Gröger, Holger Schwarz, and Bernhard Mitschang. “Prescriptive analytics for recommendation-based business process optimization”. In *International Conference on Business Information Systems*, pages 25–37. Springer, 2014.
- [4] Clyde Holsapple, Anita Lee-Post, and Ram Pakath. “A unified foundation for business analytics. *Decision Support Systems*”, 64:130{141, 2014.
- [5] Mandeep Kaur Saggi and Sushma Jain. “A survey towards an integration of big data analytics to big insights for value-creation”. *Information Processing & Management*, 54(5):758{790, 2018.
- [6] S. Kurgalin and S. Borzunov, “A Practical Approach to High-Performance Computing”. Springer. 2019.
- [7] Michael Minelli, Michele Chambers, and Ambiga Dhiraj. “Big data, big analytics: emerging business intelligence and analytic trends for today’s businesses”. Volume 578. John Wiley & Sons, 2013.
- [8] Albert Bifet and Jesse Read. “Ubiquitous artificial intelligence and dynamic data streams”. In *Proceedings of the 12th ACM International Conference on Distributed and Event-Based Systems, DEBS ’18*, Pp. 1–6, Association for Computing Machinery. New York, USA, 2018.
- [9] T. Olanrewaju. “The rise of the digital “bank.” <https://www.mckinsey.com/business-functions/digital-mckinsey/ourinsights/the-rise-of-the-digital-bank/>. Accessed: 2020.
- [10] Peter Pacheco. “An Introduction to Parallel Programming”, 1st ed., San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2011.
- [11] Federico Pigni, Gabriele Piccoli, and Richard Watson. “Digital data streams: Creating value from the real-time flow of big data”. *California Management Review*, 58(3):5–25, 2016.
- [12] Sergio Ramírez-Gallego, Bartosz Krawczyk, Salvador García, Michal Wozniak, and Francisco Herrera. “A survey on data preprocessing for data stream mining: Current status and future directions”. *Neurocomputing*, 239:39 – 57, 2017.
- [13] Stuart J Russell and Peter Norvig. “*Inteligencia Artificial: un enfoque moderno*”. 2004.
- [14] Nazmul Siddique and Hojjat Adeli. “Computational intelligence: synergies of fuzzy logic, neural networks and evolutionary computing”. John Wiley & Sons, 2013.
- [15] Usarat Thirathon, Bernhard Wieder, Zoltan Matolcsy, and Maria-Luise Ossimitz. “Impact of big data analytics

- on decision making and performance”. In International Conference on Enterprise Systems, Accounting and Logistics, 2017.
- [16] M. Turner, D. Budgen and P. Brereton, “Turning software into a service”. In Computer, vol. 36, no. 10, pp. 38-44, Oct. 2003.
- [17] Lotfi A. Zadeh. “Fuzzy logic, neural networks, and soft computing”. Commun. ACM 37, 3 , 77–84. DOI: <https://doi.org/10.1145/175247.175255>. March 1994.