DOI: https://doi.org/10.33414/ajea.1119.2022

Analítica Prescriptiva en VRP mediante Aprendizaje por Refuerzo y Flujos de Eventos

Prescriptive Analytics on VRP through Reinforcement Learning and Event Streams

Presentación: 4 y 5 de Octubre de 2022

Doctorando:
Esteban Alejandro Schab
Grupo de Investigación en Inteligencia Computacional e Ingeniería de Software, Facultad Regional Concepción del Uruguay, Universidad Tecnológica Nacional - Argentina
schabe@frcu.utn.edu.ar

Directora:
María Fabiana Piccoli

Co-director:
Carlos Antonio Casanova Pietroboni

Resumen

Los procesos de negocio exigen tomar decisiones rápidas para lograr la adaptación constante a los cambios en búsqueda de mejorar el desempeño y aprovechar las oportunidades. Resulta clave contar con analíticas que transformen los datos en conocimiento para la toma de decisiones. En este trabajo se introduce una línea de investigación enfocada en la analítica prescriptiva, capaz de calcular acciones a ser ejecutadas en el momento (decisiones operativas) o en el futuro (decisiones tácticas y/o estratégicas) para lograr un objetivo deseado, en problemas de enrutamiento de vehículos (VRP), y se presentan los avances y resultados obtenidos. El cálculo de las acciones involucra el procesamiento del flujo de eventos del negocio en forma de *datastreams*, la aplicación de técnicas y algoritmos de *Soft Computing* e *Inteligencia Computacional* (en particular *Aprendizaje por Refuerzo*) y, derivado de la necesidad de bajos tiempos de respuesta, el empleo de *Computación de Alto Desempeño*.

Palabras clave: Inteligencia Computacional, Analíticas, Aprendizaje por Refuerzo, VRP, Datastreams, Computación de Alto Desempeño.

Abstract

Business processes require quick decisions to constantly adapt to changes in order to improve performance and take advantage of opportunities. It is essential to have analytics that transform data into knowledge for decision making. This paper introduces a line of research focused on prescriptive analytics, capable of calculating actions to be executed at the moment (operational decisions) or in the future (tactical and/or strategic decisions) to achieve a desired goal, in vehicle routing problems (VRP), and presents the progress and results obtained. The calculation of the actions involves the processing of the flow of business events in the form of *datastreams*, the application of *Soft Computing* and *Computational Intelligence* techniques and algorithms (in particular *Reinforcement Learning*) and, derived from the need for low response times, the use of *High Performance Computing*.

Keywords: Computational Intelligence, Analytics, Reinforcement Learning, VRP, Datastreams, High Performance Computing.

Introducción

La mejora continua y adaptativa de los procesos de negocio es clave para las organizaciones que pretenden mantener o mejorar su competitividad. La digitalización de los procesos y el incremento en el uso de tecnologías de monitoreo, dan lugar a la producción de una enorme cantidad de datos. Estos tienen un gran potencial para la mejora de los procesos conducida por analíticas (Gröger, Schwarz, & Mitschang, 2014) (Saggi & Jain, 2018) (Thirathon, Wieder, Matolcsy, & Ossimitz, 2017). Las analíticas buscan transformar los datos en conocimiento para la toma de decisiones (Holsapple, Lee-Post, & Pakath, 2014). Pueden distinguirse cuatro tipos de analítica según el nivel de automatización del proceso (Minelli, Chambers, & Dhiraj, 2013). En orden encontramos:

- Analítica descriptiva: intenta responder qué ha pasado o está pasando.
- Analítica diagnóstica: señala por qué ha pasado o por qué está pasando.
- Analítica predictiva: aplica el conocimiento obtenido a partir de los datos para predecir nuevos datos sobre el presente o el futuro (pronóstico).
- Analítica prescriptiva: responde qué debería hacerse y cómo para lograr un objetivo. Calcula las acciones a ejecutar en el momento (decisiones operativas) o en el futuro (decisiones tácticas: corto y medio plazo, o estratégicas: largo plazo).

Los tres primeros enfoques no sugieren acciones concretas a partir de los pronósticos obtenidos, sino que descansan en el juicio subjetivo y las habilidades analíticas del usuario para deducir acciones de mejora. La analítica descriptiva y diagnóstica se basan en datos históricos.

En la actualidad, a pesar de los avances tecnológicos, la mayoría de las analíticas de procesos existentes dentro de la industria no aprovechan al máximo el conocimiento oculto en los grandes volúmenes de datos con los que cuentan (Gröger, Schwarz, & Mitschang, 2014) debido a las siguientes limitaciones:

1. Falta de uso de técnicas prescriptivas para transformar los resultados del análisis en acciones concretas de mejora. Este paso se deja a criterio subjetivo del usuario.
2. Uso intensivo de datos de los sistemas en producción. Esto implica una pérdida de rendimiento de las herramientas de software que intervienen en los procesos.
3. Por lo general, las tareas de optimización se realizan a posteriori, cuando el proceso ha finalizado. En contraste a una mejora proactiva durante la ejecución del proceso.

Debido a estas limitaciones, especialmente 2., el procesamiento de flujos de datos o Data Stream Mining (DSM) se ha convertido en un tema emergente dentro del área de Big Data (Bifet & Read, 2018) (Ramírez-Gallego, Krawczyk, García, Wozniak, & Herrera, 2017). Un datastream es una representación digital y transmisión continua de datos, los cuales describen una clase de eventos relacionada (Pigni, Piccoli, & Watson, 2016). Mediante su procesamiento, es posible lograr la toma de decisiones en tiempo real, es decir, cuando se producen los acontecimientos. Esto abre nuevas y amplias oportunidades de creación de valor en las organizaciones. Ejemplos de estos sistemas son los bancos, hospitales y comercios, con sus sistemas de atención al público; así como el funcionamiento de las Smart Grids o las aplicaciones agrícolas mediadas por sensores. En algunas organizaciones, los datos se almacenan en sus sistemas de gestión que suelen utilizar modelos relacionales de bases de datos. Estos modelos permiten la elaboración de analíticas, pero su implementación puede afectar el rendimiento, especialmente en contextos de Software as a Service (SaaS) (Turner, Budgen, & Brereton, 2003), debido a las continuas consultas necesarias para realizar la monitorización. También hay organizaciones en las que los datos se generan de forma distribuida a través de diferentes dispositivos: sensores, estaciones meteorológicas o dispositivos GPS, sin un sistema de gestión centralizado. Por estas razones, la generación de eventos y su procesamiento como flujos de datos puede constituir la tecnología de base para permitir una monitorización y toma de decisiones eficiente. En los primeros sistemas, sería un componente paralelo al sistema de atención, y en los segundos sería el propio sistema de procesamiento.

En particular, la tesis en desarrollo se centra en el problema de enrutamiento de vehículos (VRP) (Clarke & Wright, 1964) (Asghari & Mirzapour Al-e-hashem, 2021) con suministro de información en tiempo real y re-enrutamiento, orientado a la búsqueda de un paradigma de movilidad inteligente (Melo, Macedo, & Baptista, 2017). Propone el desarrollo de un modelo de analítica prescriptiva que supere los inconvenientes descritos. Este modelo dirigido por los datos será 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. El problema particular tiene las siguientes características: entrega total o parcial de productos homogéneos, restricción de la capacidad de los vehículos y demanda incierta. La solución propone el uso de agentes con aprendizaje de refuerzo con cómputo paralelo.

La construcción del modelo prescriptivo, cuya principal función resulta en la determinación de las acciones a llevar a cabo, hace uso de un modelo predictivo para explorar los futuros cercanos y un modelo descriptivo para calcular la aptitud de dichos estados. Para ello se propone el uso de agentes basados en aprendizaje por refuerzo, y complementar el mismo con técnicas provenientes de la Inteligencia Computacional: redes neuronales como modelos, teoría de conjuntos difusos como lenguaje de especificación, y métodos numéricos y metaheurísticos para el entrenamiento de tales modelos (Ebrahimnejad & Verdegay, 2018) (Siddique & Adeli, 2013) (Zadeh, 1994). Debido a la necesidad de dar respuesta a un proceso de negocio dinámico, la ejecución de estas técnicas y algoritmos debe ser lo suficientemente rápida como para procesar los *datastreams* que el sistema genera de manera continua y brindar resultados en tiempo real, lo que implica el uso de técnicas y herramientas HPC (Kurgalin & Borzunov, 2019).

Desarrollo

El problema de enrutamiento de vehículos (VRP) es una versión más general del TSP (Travelling Salesman Problem) (Flood, 1956). Su principal diferencia es considerar múltiples vehículos en su modelo de enrutamiento. Es decir, hay un conjunto de clientes, dispersos geográficamente alrededor de un depósito central y una flota de vehículos homogénea (Clarke & Wright, 1964) (Asghari & Mirzapour Al-e-hashem, 2021). El VRP trata de cómo servir de forma óptima a todos sus clientes. Desde su formulación en 1959, la modelización de VRP ha sido uno de los temas más abordados en el marco de la investigación operativa, la ingeniería industrial, la logística y el transporte. En particular, este trabajo considera una variante de VRP con las siguientes características:

- Bienes homogéneos: los bienes o mercancías distribuidos son homogéneos y divisibles. Las demandas deben satisfacerse en la medida de lo posible para la mayoría de los clientes.
- Capacidad de los vehículos: los vehículos tienen una capacidad que no puede superarse en cada viaje.
- Demanda incierta: los clientes están distribuidos en la red y sus demandas no se conocen de antemano, aunque se puede disponer de estimaciones provenientes de analíticas predictivas.
- Condiciones dinámicas de la red: durante la etapa de operación, los tiempos de viaje entre los lugares de entrega o reabastecimiento pueden cambiar debido a diversos eventos. Ejemplos de ellos son los atascos, los cortes de calles por mantenimiento y las movilizaciones de personas, entre otros. Estos cambios no suelen conocerse de antemano, es posible percibir el estado de la red en cualquier momento y actualizar todos los parámetros.
- Cartera de clientes estable: los clientes y su ubicación rara vez varían con el tiempo.

Por lo tanto, nuestra variante VRP tiene una capacidad vehicular (C) y condiciones de incertidumbre (U), es dinámica (D) y en tiempo real (RT). Por lo que la llamamos RT-CUD-VRP.

El aprendizaje es un proceso en el que los parámetros libres de un modelo se adaptan a través de la estimulación recibida del entorno en el que está inmerso. El tipo de aprendizaje viene determinado por la forma en que se producen los cambios en los parámetros (Haykin, 1994). Un conjunto de reglas bien definidas para la solución del problema de aprendizaje se denomina algoritmo de aprendizaje o método de entrenamiento. No existe un único algoritmo de entrenamiento, sino una gran variedad, y cada uno tiene sus propias ventajas e inconvenientes. Una vez que un modelo ha sido entrenado por algún algoritmo de aprendizaje y se han establecido sus parámetros libres, se dice que el modelo ha

aprendido y puede realizar las tareas para las que fue entrenado sin que se alteren sus parámetros. El tipo de retroalimentación suele ser el factor más importante a la hora de determinar la naturaleza del problema de aprendizaje abordado. Se distinguen tres tipos de aprendizaje: supervisado, no supervisado y por refuerzo.

El aprendizaje por refuerzo (RL) es el más relevante en este contexto. Un agente es capaz de juzgar y criticar sus acciones teniendo en cuenta sus percepciones y alguna medida de aptitud, recompensa o refuerzo. La tarea de aprendizaje por refuerzo consiste en utilizar las recompensas observadas para aprender una política óptima (o casi óptima) del entorno (Russell & Norvig, 2004), es decir, la que maximiza las recompensas totales recibidas de su interacción con el contexto. Esta política indica al agente qué hacer en cada estado posible.

En un sistema de aprendizaje por refuerzo, es posible identificar cuatro subelementos principales: una política, una señal de recompensa, una función de valor y, opcionalmente, un modelo de entorno. En (Schab, Casanova, & Piccoli, 2022b) se describe con detalle cada uno de estos elementos y sus interacciones.

Los métodos que aprenden aproximaciones tanto a la política como a las funciones de valor suelen llamarse métodos *actor-crítico*. La política está parametrizada, y el algoritmo de aprendizaje las ajusta de acuerdo con una regla de ascenso de gradiente estocástico mientras acumula experiencia al interactuar con el entorno y obtener observaciones y recompensas (Sutton & Barto, 2018).

La tesis en desarrollo propone un modelo de analítica prescriptiva para RT-CUD-VRP mediante el uso de agentes de aprendizaje por refuerzo aplicando técnicas de computación paralela en GPU (Graphics Processing Unit) para acelerar el entrenamiento. Dadas las características de RT-CUD-VRP, se seleccionó un agente Actor-Crítico utilizando el método de Optimización de Políticas Proximales (PPO) (Schulman, Dhariwal, Radford, & Klimov, 2017). Este algoritmo alterna entre la generación de experiencia (almacenada como trayectorias, a través de la interacción con el entorno) y la optimización de la función objetivo (por ascenso de gradiente estocástico utilizando la experiencia generada). Esta forma de trabajar, al generar varias épocas de actualizaciones en mini lotes, permite acelerar el entrenamiento utilizando computación de alto rendimiento (HPC).

Como describen (Sutton & Barto, 2018) y (Barto, Sutton, & Anderson, 2021), los métodos de aprendizaje por refuerzo se han aplicado hasta ahora a problemas sencillos de toma de decisiones, principalmente relacionados con la resolución de juegos, a estados representados con matrices de dimensiones fijas procedentes del procesamiento de imágenes o sensores, y a decisiones básicas. El gran reto por delante es adaptar los métodos actor-crítico para abordar una gama cada vez más amplia de problemas del mundo real de importancia científica y social. RL tiene el potencial de mejorar la calidad, la eficiencia y la rentabilidad de los procesos de los que dependemos en la educación, la sanidad, el transporte y la gestión de la energía, entre otros. Para conseguirlo, hay que abordar las decisiones de diseño y los ajustes que implica la implantación del RL. Hay que diseñar la arquitectura seleccionando los algoritmos de aprendizaje adecuados, las representaciones de estados y acciones, los procedimientos de entrenamiento, la configuración de los hiperparámetros y otros detalles de diseño (Barto, Sutton, & Anderson, 2021).

Para resolver RT-CUD-VRP utilizando el algoritmo PPO, diseñamos el entorno, las observaciones, las acciones, las recompensas, y la función de valor y la política con sus algoritmos de entrenamiento. En (Schab, Casanova, & Piccoli, 2022b) se describen cada uno de ellos.

La implementación se desarrolló con las APIs de TensorFlow, principalmente TF-Agents (Hafner, Davidson, & Vanhoucke, 2017), el cuál proporciona un paradigma para el diseño, implementación y testeo de agentes de aprendizaje por refuerzo en paralelo. Para la integración de RT-CUD-VRP en un entorno dinámico y complejo, se desarrolló una simulación mediante la librería Simpy. El código desarrollado está disponible en (Schab, Casanova, & Piccoli, 2022a).

Se diseñaron varias instancias como forma de validación inicial y con el objetivo de mostrar los puntos fuertes y débiles del modelo. La instancia base está conformada por 9 clientes y un único depósito. Las demandas de cada cliente se modelan mediante números difusos trapezoidales, y los tiempos de descarga mediante números difusos triangulares. Los tiempos de viaje se consideraron estáticos en una instancia y dinámicos en las otras. A su vez, el muestreo de los números difusos genera instancias con baja y alta incertidumbre en las demandas y tiempos de descarga. Para hacer comparaciones justas, se utilizó la misma arquitectura para el agente en todas las simulaciones, y la misma plataforma de ejecución de los experimentos. Se pueden observar los resultados iniciales en (Schab, Casanova, & Piccoli, 2022b).

Conclusiones

En este trabajo, se describen la motivación y los conceptos principales de la tesis en desarrollo, y se presentan los avances y resultados obtenidos. La tesis propone un modelo prescriptivo para el Problema de Enrutamiento de Vehículos: RT-CUD-VRP. Sus características son entrega total o parcial de mercancías homogéneas, restricción de la capacidad de los vehículos, condiciones de incertidumbre, dinámica y tiempo real. La propuesta utiliza agentes de aprendizaje por refuerzo con cómputo paralelo en GPU. Se mencionan los principales detalles de implementación en código abierto.

Los primeros resultados obtenidos muestran que el agente puede aprender una política que funciona suficientemente bien en tres casos diseñados, mostrando los puntos fuertes y débiles del modelo. Los resultados iniciales son alentadores y motivan la realización de nuevos trabajos. Como líneas futuras, se espera desarrollar un análisis completo de la propuesta mediante una comprobación experimental exhaustiva. También se pretende extender el modelo para resolver cualquier variante de VRP dinámico en tiempo real. Dicha generalización requerirá una representación universal de los estados, debiendo explorar técnicas de Graph Embeddings o Graph Encoding. Por último, queremos validar nuestra solución en otros problemas con la misma naturaleza de VRP. Respecto al rendimiento computacional, tenemos que hacer una evaluación comparativa con otras soluciones, con o sin HPC.

Referencias

Asghari, M., & Mirzapour Al-e-hashem, S. (2021). Green vehicle routing problem: A state-of-the-art review. *International Journal of Production Economics*, 107899. doi:https://doi.org/10.1016/j.ijpe.2020.107899

Barto, A. G., Sutton, R. S., & Anderson, C. W. (2021). Looking Back on the Actor–Critic Architecture. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 40-50. doi:10.1109/TSMC.2020.3041775

Bifet, A., & Read, J. (2018). Ubiquitous artificial intelligence and dynamic data streams. *Proceedings of the 12th ACM International Conferenceon Distributed and Event-Based Systems, DEBS '18* (págs. 1–6). New York, USA: Association for Computing Machinery.

Clarke, G., & Wright, J. W. (1964). Scheduling of Vehicles from a Central Depot to a Number of Delivery Points. *Operations Research, 12*(4), 568-581. Obtenido de http://www.jstor.org/stable/167703

Ebrahimnejad, A., & Verdegay, J. L. (2018). *Fuzzy sets-based methods and techniques for modern analytics.* Springer International Publishing.

Flood, M. M. (1956). The Traveling-Salesman Problem. *Operations Research*, 61-75.

Gröger, C., Schwarz, H., & Mitschang, B. (2014). Prescriptive analytics for recommendation-based business process optimization. *International Conference on Business Information Systems*, 25–37.

Hafner, D., Davidson, J., & Vanhoucke, V. (2017). TensorFlow Agents: Efficient Batched Reinforcement Learning in TensorFlow. *CoRR*. Obtenido de http://arxiv.org/abs/1709.02878

Haykin, S. (1994). *Neural Networks: a Comprehensive Foundation.* NY: Macmillan.

Holsapple, C., Lee-Post, A., & Pakath, R. (2014). A unified foundation for business analytics. *Decision Support Systems*(64), 130-141.

Kurgalin, S., & Borzunov, S. (2019). *A Practical Approach to High-Performance Computing.* Springer.

Melo, S., Macedo, J., & Baptista, P. (2017). Guiding cities to pursue a smart mobility paradigm: An example from vehicle routing guidance and its traffic and operational effects. *Research in transportation economics, 65*, 24-33.

Minelli, M., Chambers, M., & Dhiraj, A. (2013). *Big data, big analytics: emerging business intelligence and analytic trends for today's businesses.* John Wiley & Sons.

Pigni, F., Piccoli, G., & Watson, R. (2016). Digital data streams: Creating value from the real-time flow of big data. *California Management Review, 58*(3), 5–25.

Ramírez-Gallego, S., Krawczyk, B., García, S., Wozniak, M., & Herrera, F. (2017). A survey on data preprocessing for data stream mining: Current status and future directions. *Neurocomputing, 239*, 39 – 57.

Russell, S., & Norvig, P. (2004). *Inteligencia Artificial: un enfoque moderno.* Pearson Prentice Hall.

Saggi, M. K., & Jain, S. (2018). A survey towards an integration of big data analytics to big insights for value-creation. *Information Processing & Management, 54*(5), 758-790.

Schab, E. A., Casanova, C. A., & Piccoli, M. F. (Abril de 2022). *Reinforcement Learning for VRP*. Obtenido de GitHub repository: https://github.com/estebanschab/RL-VRP

Schab, E. A., Casanova, C. A., & Piccoli, M. F. (2022). Solving an Instance of a Routing Problem Through Reinforcement Learning and High Performance Computing. *Conference on Cloud Computing, Big Data & Emerging Topics* (págs. 107-121). La Plata: Springer, Cham.

Schulman, J. a., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.

Siddique, N., & Adeli, H. (2013). *Computational intelligence: synergies of fuzzy logic, neural networks and evolutionary computing.* John Wiley & Sons.

Sutton, R., & Barto, A. (2018). *Reinforcement learning: An introduction.* MIT press.

Thirathon, U., Wieder, B., Matolcsy, Z., & Ossimitz, M.-L. (2017). Impact of big data analytics on decision making and performance. *International Conference on Enterprise Systems, Accounting and Logistics.*

Turner, M., Budgen, D., & Brereton, P. (2003). Turning software into a service. *Computer, 36*(10), 38-44.

Zadeh, L. (1994). Fuzzy logic, neural networks, and soft computing. *Communications of the ACM, 37*(3), 77-84. doi:10.1145/175247.175255