

Metaheurísticas, búsqueda estocástica y cómputo eficiente en optimización aplicada

Tomás Tetzlaff¹, Adriana Gaudiani¹, Andrés Rojas Paredes¹, Diego Encinas², Esteban Fassio¹, Mariano Trigila³, Rodrigo González¹, Daniel Bertaccini¹

¹Instituto de Ciencias, Universidad Nacional de General Sarmiento

²Proyecto SimHPC - Programa TICAPPS - Instituto de Ingeniería y Agronomía - Universidad Nacional Arturo Jauretche

³Facultad de Ingeniería y Ciencias Agrarias, Pontificia Universidad Católica Argentina-CABA

{tetzlaff, agaudiani, arojas, efassio, dbertacc, rgonzalez}@campus.ungs.edu.ar, dencinas@unaj.edu.ar, mariano_trigila@uca.edu.ar

Resumen

Las metaheurísticas son técnicas de optimización y resolución de problemas computacionales que toman inicialmente una solución factible, la cual es luego mejorada usando procedimientos heurísticos conocidos, como recocido simulado, algoritmos genéticos, búsqueda tabú y redes neuronales. La búsqueda estocástica está presente en estos métodos y su importancia reside en ser una herramienta general de optimización cuyo estudio puede aportar mejoras para las metaheurísticas y desarrollar variantes de ellas. Este proyecto propone estudiar propiedades teóricas y prácticas de estas técnicas y su aplicación en los problemas en los que trabajan sus integrantes, brindando metodologías para incrementar la eficiencia de los algoritmos involucrados y la confiabilidad de los resultados que producen. Estas metodologías aprovechan los avances en los métodos y técnicas de la computación eficiente y del cómputo en paralelo para desarrollar los algoritmos necesarios para validar las nuevas propuestas mediante casos de experimentación. La implementación de los algoritmos que permiten realizar las experiencias requiere la utilización de técnicas de las ciencias de los datos, del cómputo de alto rendimiento y métodos del campo de la optimización.

Palabras Claves: *metaheurísticas, técnicas de optimización, estimación de parámetros, soluciones factibles, eficiencia.*

Contexto

La línea de Investigación que se presenta corresponde al Proyecto: 30/1147 - Metaheurísticas, búsqueda estocástica y aplicaciones con cómputo eficiente, del Instituto de Ciencias de la Universidad Nacional de General Sarmiento (UNGS). (2020-2022).

Existe cooperación con grupos de investigación en otras Universidades Argentinas y del exterior, con las cuales los integrantes del proyecto han participado en investigaciones en conjunto.

Introducción

Un método heurístico es un procedimiento que, teniendo conocimiento de un problema y de las técnicas aplicables, aporta soluciones o se acerca a ellas usando una cantidad de recursos razonable. Por lo general, este recurso es el tiempo de cómputo que se emplea en la búsqueda de la solución. Se trata de procedimientos inteligentes en el sentido de realizar una tarea que no es producto de un riguroso análisis formal ya conocido y aplicable al tema, que es a veces ineficiente en cuanto al tiempo de cómputo, sino de algoritmos de simulación, de búsqueda o evolutivos que utilizan y también aportan al conocimiento experto sobre la tarea. Una solución heurística de un problema es aquella proporcionada por un método heurístico, es decir, aquella solución sobre la que se tiene cierta confianza de que alcanza algún grado de optimalidad y/o factibilidad [1].

Las metaheurísticas (también llamadas heurísticas modernas) se han desarrollado principalmente a partir de la década de 1980. En general, las metaheurísticas toman inicialmente una solución factible, la cual es luego mejorada usando heurísticas de mejoramiento embebidas en una estructura más general, i.e., Recocido Simulado (*Simulated Annealing*, SA), Algoritmos Genéticos (*Genetic Algorithms*, GA), Búsqueda Tabú (*Tabu Search*, TS), y Redes Neuronales (*Neural Networks*, NN). La característica común de estos enfoques es el uso de mecanismos para encontrar en lo posible soluciones óptimas evadiendo óptimos locales [2].

Las metaheurísticas pueden manejar variables discretas y reales, siendo aplicadas en la actualidad a una amplia gama de problemas de optimización de manera efectiva. Básicamente, los enfoques metaheurísticos basados tanto en la trayectoria (sucesión de posibles soluciones) como en la población (conjunto adaptable de posibles soluciones) apuntan a ubicar el óptimo global en el espacio de la solución a través de movimientos aleatorios. La diferencia clave entre las metaheurísticas está en la forma en que proponen el siguiente movimiento en el espacio de la solución [3]

El avance constante de las tecnologías de cómputo provee recursos computacionales con más rapidez de procesamiento, lo cual permite a investigadores y profesionales de la ciencia y la ingeniería resolver problemas combinatorios complejos y de gran escala. Esto genera un interés creciente en resolver problemas complejos del mundo real utilizando métodos metaheurísticos implementados con técnicas que mejoran la eficiencia, como el cómputo paralelo que permite reducir los tiempos de procesamiento y obtener la mayor eficiencia de las plataformas de hardware heterogéneas mediante el diseño y desarrollo de algoritmos paralelos adecuados para lograrlo [4].

En particular la Ingeniería del Software, a pesar de ser un área reciente, es en la actualidad una importante fuente de problemas de optimización. Los ingenieros y gestores de proyectos de software se enfrentan diariamente a diversos problemas de optimización en los que las técnicas exactas no tienen cabida por el corto intervalo de tiempo en que se requiere una respuesta [5]. En Ingeniería de requerimientos el procesamiento de texto y la extracción automática de contenido es un tema en investigación en rápido desarrollo [6]. Durante el preprocesado del texto se obtienen características decisivas para la clasificación y priorización de contenidos. Luego se aplican técnicas avanzadas de procesamiento y análisis de lenguaje natural [7] [8]. Las técnicas heurísticas ayudan a la estimación de parámetros de los métodos predictivos que permiten acceder al conocimiento latente en el texto analizado.

Cabe la posibilidad, por tanto, de aplicar a estos problemas, algoritmos metaheurísticos que ofrezcan al ingeniero una solución de cierta calidad en un breve periodo de tiempo: un compromiso entre calidad de la solución y la rapidez en encontrarla. En éstos y otros campos de investigación a menudo es esencial modelar y resolver tareas de optimización, de aprendizaje o de investigación para aplicaciones que no admiten una fácil formulación.

Líneas de Investigación y Desarrollo

Se propone abordar las siguientes líneas de I+D desde la perspectiva del estudio de las metaheurísticas y del uso del cómputo de alto rendimiento (HPC), ambos como ejes centrales:

- Evaluación de escalabilidad y rendimiento mediante balance de carga, mapping y scheduling de procesos, métricas de prestación y de consumo.
- Paralelización de aplicaciones (numéricas y no numéricas) con alta demanda computacional y/o grandes volúmenes de

datos sobre arquitecturas multiprocesador distribuidas (puras e híbridas).

- Desarrollo de soluciones aplicadas a problemas de impacto social con requerimiento de HPC.
- Aplicaciones de la teoría de probabilidad. Propiedades y simulación de procesos estocásticos. Búsqueda estocástica. Metaheurísticas.
- Estadística matemática y sus aplicaciones en espacios de búsqueda. Métodos de Monte Carlo.

Algunas de los estudios que se abordan en este trabajo ya vienen desarrollándose en la UNGS y en los grupos externos que colaboran con este proyecto, otros se inician a partir de cooperaciones con otras Universidades o para formalizar trabajos finales de grado o postgrado.

Se ha iniciado una cooperación en el área de Simulación para investigación y cursos de postgrado con el Proyecto SimHPC de la Universidad Nacional Arturo Jauretche y para formalizar trabajos finales. Entre otros temas el Proyecto SimHPC enfoca al desarrollo de modelos y simuladores para el área de Salud como así también para el tratamiento de grandes volúmenes de datos.

Existe una fuerte colaboración, con el grupo HPC4EAS (*High Performance Computing for Efficient Applications and Simulation*) del Departamento de Arquitectura de Computadores y Sistemas Operativos de la Universidad Autónoma de Barcelona en la dirección de tesis de postgrado. Se espera consolidar a corto plazo un grupo de investigación, desarrollo y transferencia que generará oportunidades de formación de recursos humanos, proveerá de servicios a la comunidad en el área de estudio y potenciará los vínculos de cooperación con otras instituciones.

Resultados y Objetivos

Este proyecto tiene como objetivo principal el estudio y análisis de la aplicación de metaheurísticas y técnicas de búsqueda estocástica, que permitan optimizar el funcionamiento de programas utilizados en diversas áreas de desarrollo. Es también un objetivo en todos los problemas de aplicación incrementar la eficiencia de los algoritmos involucrados y la confiabilidad de los resultados que producen. En particular nos referiremos a los siguientes problemas:

- **Optimización de la simulación de sistemas físicos** En la Ingeniería Hidrológica, una amplia gama de aplicaciones con diferentes y novedosas metodologías y técnicas metaheurísticas son creadas para abordar una variedad de problemas del modelado hidrológico [9] [10]. Se busca mejorar la simulación mediante métodos de búsqueda de un conjunto óptimo de parámetros del modelo, interpretando como óptimo a la mejor aproximación a la realidad que se puede hacer con ese modelo. Se continúan los trabajos de investigación en curso aplicando heurísticas para estimar los parámetros de ajuste que disminuyen la complejidad y la probabilidad de errores en un simulador del cauce del Río Paraná y contribuir a mejorar su predicción con el menor costo computacional posible [11] [12][13].
- **Métodos de aceleración del procesamiento de lenguaje natural.** Los usuarios de aplicaciones móviles generan a través de las *app stores* comentarios y calificaciones para describir su experiencia con las aplicaciones que descargan y usan. Estos comentarios, también llamados *reviews*, tienen información muy valiosa para retroalimentar el proceso de desarrollo de *apps*, por ejemplo, se pueden encontrar comentarios en forma de pedidos de nuevos features y reporte de errores. Este tipo de información es relevante para el desarrollo y la evolución de las aplicaciones móviles [14]. La enorme expansión de las aplicaciones móviles y el constante crecimiento de la cantidad de *reviews* de usuarios provocan un cuello de botella que surge del procesamiento de

grandes volúmenes de datos. Se espera aplicar metaheurísticas para encontrar de manera eficiente los parámetros de los métodos predictivos que buscan el conocimiento latente en el texto. Se propone crear un pipeline de preprocesamiento de las *reviews* en modo secuencial y compararlo contra una implementación con HPC. Se aplicarán las técnicas clásicas de paralelismo: paralelismo de datos o paralelismo de tareas, y combinación de ambas técnicas. Se propone explorar diferentes tecnologías como multicore, GPGPU, cluster standard y cluster de Raspberry. Adicionalmente nos preguntamos sobre la eficiencia energética de nuestra solución. En este punto nos proponemos realizar una implementación sobre un cluster de Raspberry que es la alternativa más adecuada en términos energéticos. Esperamos cuantificar el gasto energético de nuestra solución.

- **Fundamentos teóricos de la búsqueda estocástica:** La búsqueda estocástica permite encontrar soluciones satisfactorias a un problema en algunos casos en tiempos muy reducidos, aun careciendo de la fundamentación sobre porqué se llegó a las mismas. Nos proponemos seguir avanzando con la teoría matemática que fundamenta las propiedades y la eficiencia de estas metodologías [15][16]. Asimismo, esperamos realizar aportes rigurosos en este campo para los métodos usados en aplicaciones biológicas [17].

Se espera consolidar un grupo de investigación, desarrollo y transferencia a la comunidad con capacidad para abordar estos tipos de problemas.

Formación de Recursos Humanos

Tres miembros del equipo poseen formación de postgrado a nivel de doctorado, uno de ellos es especialista en Cómputo de Altas Prestaciones. Uno de los investigadores está en su etapa final para obtener el grado de Doctor en Ciencias Informáticas, dirigido por uno de los miembros del proyecto y un investigador de la UAB

(España). Otro investigador está comenzando una Maestría en Ciencia de los Datos, de interés para este proyecto. Un área afin a los del proyecto (evolución por deriva estocástica en poblaciones biológicas) es el tema de una tesis en curso (Doctorado en Ciencia y Tecnología, UNGS), dirigida por un integrante del proyecto. Uno de los integrantes coordina una Diplomatura en Ciencias de Datos.

Bibliografía

1. Reeves C.R. Fitness Landscapes. In: Burke E., Kendall G. (eds) Search Methodologies, pp 681-705. Springer, Boston, MA. (2014) https://doi.org/10.1007/978-1-4614-6940-7_22
2. Du K.L., Swamy M.N.S.: Genetic Algorithms. In: Search and Optimization by Metaheuristics. Birkhäuser, Cham, (2016)
3. Piotrowski, A.P., Napiorkowski, J.J. Some metaheuristics should be simplified. Inf. Sci. (Ny) **427**, 32–62, (2018)
4. H Hennessy, J. L., Patterson, and D. A., Computer Architecture, Fourth Edition: A Quantitative Approach. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. (2006)
5. Tekiner F., Tsuruoka Y., Tsujii J., Ananiadou S., Keane J. Parallel Text Mining for Large Text Processing, pp. 348-353 In: Proceedings of IEEE CSNDSP 2010, 21-23 July, Newcastle, UK (2010)
6. E. Guzman, R. Alkadhi, and N. Seyff, “A Needle in a Haystack: What Do Twitter Users Say about Software? In: Proc. of the International Requirements Engineering Conference, pp. 96–105. (2016)
7. E. Oehri and E. Guzman. Same Same but Different: Finding Similar User Feedback Across Multiple Platforms and Languages, 2020 IEEE RE, Zurich, Switzerland, pp. 44-54. (2020) doi: 10.1109/RE48521.2020.00017
8. Guzman E. and W. Maalej. How do users like this feature? A fine-grained sentiment analysis of app reviews. In; 22nd IEEE International Requirements Engineering Conference (RE'14), p. 153–162, (2014).
9. Chau K.-W., Use of meta-heuristic techniques in rainfall-runoff modelling, *Water*, vol. 9, no. 3, pp. 186, (2017)

10. Fallah-Mehdipour E., Haddad O.B.: Application of Genetic Programming in Hydrology. In: Gandomi A., Alavi A., Ryan C. (eds) Handbook of Genetic Programming Applications. Springer, Cham, (2015).
11. Gaudiani, E. Luque, P. García, M. Re, A. De Giusti, M. Naiouf. Computing, a powerful tool for improving the parameters simulation quality in flood prediction. *Procedia Computer Science – Volume 29*, p. 299-309, (2014).
12. Gaudiani, E. Luque, P. García, M. Re, A. De Giusti, M. Naiouf. How a computational method can help to improve the quality of river flood prediction by simulation. In: *Advances and new Trends in Environmental Informatics*. Springer-Verlag Berlin Heidelberg – Alemania. Volume 5, p.337-351, (2015).
13. Trigila M., Gaudiani A., Luque E. (2018) Agile Tuning Method in Successive Steps for a River Flow Simulator. In: Shi Y. et al. (eds) *Computational Science – ICCS 2018, Lecture Notes in Computer Science*, Volume 10862. Springer, Cham, (2018).
14. E. Guzman and A. Paredes Rojas. Gender and User Feedback: An Exploratory Study, *IEEE 27th International Requirements Engineering Conference (RE)*, pp. 381-385. (2019) doi: 10.1109/RE.2019.00049.
15. Maffei A.L., Tetzlaff T. Propiedades de un procedimiento de búsqueda evolutiva con diversidad. LX Reunión Anual de Comunicaciones Científicas de la Unión Matemática Argentina. Tandil, Buenos Aires. (2010)
16. Maffei A.L., Tetzlaff T. A schema theorem for stochastic search with forced diversity. Seventh Regional Meeting on Probability and Mathematical Statistics. Santa Fe, Argentina, (2010).
17. Marengo J., Tetzlaff T., Cálculo recursivo de probabilidad de presencia de secuencias en cadenas markovianas, 13° Simposio Argentino de Investigación Operativa, 44 Jornadas Argentinas de Informática, Universidad Nacional de Rosario, Rosario, Argentina. (2015)