

Hiperheurística basada en minería de datos para la resolución de problemas de caja negra

Jesús Sánchez-Oro
Director: Abraham Duarte
Co-director: Rafael Martí

Dept. Ciencias de la Computación, Arquitectura de Computadores, Lenguajes y
Sistemas Informáticos, Estadística e Investigación Operativa
Universidad Rey Juan Carlos, Spain,
jesus.sanchezoro@urjc.es
Fecha de inicio de la tesis: 2013

Resumen El estudio y resolución de problemas complejos mediante el uso de metaheurísticas es un tema muy extendido en el área de la inteligencia artificial. Sin embargo, estas metaheurísticas o los algoritmos heurísticos que las componen suelen estar diseñadas para obtener la máxima calidad cuando son aplicadas sobre un problema concreto. Eso hace que, en la mayoría de los casos, un mismo algoritmo que es considerado estado del arte para un problema determinado, no sea competitivo cuando se aplica a otro problema similar. Por otra parte, en los últimos años ha aumentado el interés por los algoritmos de caja negra. En estos algoritmos, no se posee información a priori del problema que se está tratando solucionar, tan solo se dispone de una función de evaluación de la calidad de una solución. El objetivo principal de esta tesis es el desarrollo de una hiperheurística que utilice la minería de datos para extraer información de utilidad de un problema de caja negra y utilizarla para proporcionar soluciones competitivas con el estado del arte.

Keywords: hiperheurísticas, metaheurísticas, caja negra, optimización combinatoria

1. Introducción

La optimización es un área que trata de buscar soluciones factibles a problemas de la vida real y tiene aplicaciones en ingeniería, medicina, economía y otras muchas áreas científicas [10]. Desde el punto de vista matemático, un problema de optimización (OP, del inglés *Optimization Problem*) se puede definir como la minimización o maximización del valor de la función objetivo $f(x)$, ateniéndose a un conjunto de restricciones. La solución óptima a un OP es aquella con el menor (minimización) o mayor (maximización) valor de función objetivo de entre todas las soluciones factibles.

Los problemas de optimización combinatoria (COP, del inglés *Combinatorial Optimization Problem*) son un tipo de problemas de optimización cuyas

soluciones están formadas por números enteros [15]. Más formalmente [1], un $COP = (S, f)$ se podría definir como la maximización o minimización de $f(x)$ con S representando al espacio de soluciones factibles del problema. De manera resumida:

$$COP = \begin{cases} X = \{x_1, x_2, \dots, x_n\} \rightarrow f(x) \\ D_1, \dots, D_n \\ f : D_1 \times D_2 \times \dots \times D_n \\ S = \{s = \{(x_1, v_1), \dots, (x_n, v_n)\} | v_i \in D_i\} \end{cases}$$

donde X es un conjunto de variables, D_1, \dots, D_n su dominio, f una función objetivo y S un espacio de soluciones en el cual s satisface todas las restricciones. Cada variable x_i toma un valor v_i dentro del dominio $D_i \in \mathbb{N}(i = 1, \dots, n)$, de manera que la solución $s \in S$ cumple con las restricciones del problema.

2. Hipótesis y objetivos

Una vez definida el área de trabajo en la que se enmarca la Tesis Doctoral, es necesario formular la hipótesis que se establece como base de la investigación que se va a llevar a cabo, así como los distintos objetivos que se pretenden alcanzar. La optimización combinatoria es un área muy trabajada tanto desde el punto de vista heurístico como exacto, con el desarrollo de algoritmos específicos para la resolución de problemas concretos, como por ejemplo *Vertex Separation Problem* [23], *Profile Minimization Problem* [21], o *Cutwidth Minimization Problem* [14]. Sin embargo, la mayor parte de los trabajos en optimización combinatoria proponen algoritmos para resolver problemas específicos. Por eso, consideramos de gran interés para la comunidad científica el estudio de un algoritmo de caja negra, que utilice técnicas de minería de datos para extraer características del problema que tiene que resolver. La hipótesis planteada es comprobar que un algoritmo de este tipo puede llegar a ser competitivo con los algoritmos específicos propuestos para cada problema.

El principal objetivo de esta Tesis es el desarrollo de una hiperheurística que sea capaz de proporcionar resultados competitivos con los mejores algoritmos disponibles para un conjunto de problemas de optimización combinatoria, sin disponer de información a priori de cada problema. Para ello, el algoritmo propuesto hará uso de técnicas de minería de datos que le permitan extraer características de los problemas. Estas características ayudarán al algoritmo a decidir de manera automática qué metaheurística de las disponibles se adapta mejor al problema que se está tratando de resolver. Una vez elegida la metaheurística, el algoritmo la aplicará sobre el problema en concreto, y la calidad de los resultados obtenidos servirá como retro-alimentación para la hiperheurística, de manera que sirva al algoritmo para tomar futuras decisiones sobre nuevos problemas. La característica dinámica de la propuesta permite que el algoritmo mejore su aprendizaje tras cada problema al que se enfrente, lo que eventualmente le llevará a mejorar los resultados obtenidos.

El conjunto de metaheurísticas disponibles en el algoritmo principal se seleccionarán en la primera fase de la Tesis, donde se desarrollarán diferentes algoritmos para un conjunto de problemas específicos. Esta primera fase servirá para que el doctorando adquiera experiencia en el desarrollo de este tipo de algoritmos, así como para plantear posibles estrategias para la extracción de características del algoritmo final.

Es importante destacar que existen estudios previos referentes a algoritmos de caja negra para la resolución de problemas genéricos, pero no utilizan la minería de datos para mejorar los resultados. En concreto, Gortázar et al. [11] proponen un algoritmo para la resolución de problemas binarios basado en búsqueda dispersa, mientras que Laguna et al. [13] proponen una variante adaptada a problemas de variables enteras. Además, diferentes organizaciones comercializan software de optimización como CPLEX¹ o Gurobi², que utilizan métodos exactos para resolver problemas de optimización, lo que justifica el interés de esta Tesis, tanto académico como de transferencia.

3. Metodología y plan de trabajo

El proceso de investigación en optimización sintetiza los pasos básicos a seguir en el desarrollo de una investigación científica en el área. Una vez identificada la hipótesis a trabajar durante la Tesis, así como los objetivos a llevar a cabo, el proceso comienza con una revisión del estado del arte en diversos problemas de optimización combinatoria. Como resultado de dicha revisión se identifican los algoritmos desarrollados, así como el conjunto de instancias empleadas en su evaluación. A continuación, para cada problema analizado, se plantea una hipótesis para su resolución, que desembocará en el desarrollo de uno o varios algoritmos específicos.

El proceso de investigación descrito se muestra de manera gráfica en el diagrama de actividad de la Figura 1. Si los resultados finales satisfacen la hipótesis inicial, el proceso concluye con la publicación de la misma junto con los resultados experimentales obtenidos.

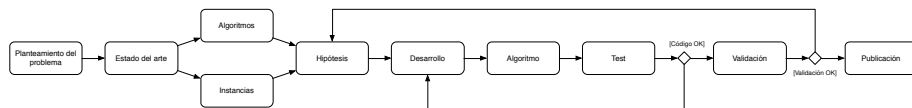


Figura 1: Método de investigación

¹ <http://www-01.ibm.com/software/commerce/optimization/cplex-optimizer/>

² <http://www.gurobi.com/>

4. Relevancia

Esta Tesis tiene una relevancia notable tanto en el ámbito de la investigación como en el de la transferencia a la sociedad. En el ámbito académico, el algoritmo propuesto será de gran utilidad para la comunidad científica, ya que podrá ser utilizado como marco base de comparación para cualquier nueva propuesta algorítmica. Además, podrá ser incluido en algoritmos exactos como método de pre-procesamiento que permita reducir el espacio de búsqueda, y, por lo tanto, mejorar la eficiencia y eficacia de este tipo de algoritmos. Esta característica puede ser de gran interés para programas comerciales como CPLEX o Gurobi.

En cuanto a la transferencia a la sociedad, cabe destacar que, al tratarse de un algoritmo de caja negra, podrá ser aplicado a todo tipo de problemas (existentes o nuevos) sin necesidad de modificaciones que conlleven una gran inversión de tiempo y dinero. Esta característica permitirá que cualquier empresa sea capaz de utilizar un único algoritmo para solucionar cualquier problema de optimización que se presente de manera inmediata, reduciendo por tanto sus costes y plazos de entrega. Por ejemplo, una empresa que utilice el algoritmo para optimizar los horarios de sus trabajadores, podrá, sin realizar grandes modificaciones, utilizar el mismo algoritmo para optimizar las rutas que siguen sus vehículos de entrega de mercancías o para optimizar la cantidad de material que se utiliza en la construcción de sus productos. Además, la naturaleza dinámica del algoritmo hace que su calidad mejore con el incremento del número de problemas a los que se enfrenta.

Los resultados parciales obtenidos por la Tesis en desarrollo remarcen la relevancia del proyecto. En concreto, se han publicado 9 artículos en revistas indexadas en el *Journal of Citation Reports* (JCR) [2,4,5,6,7,20,21,23,22]. De estos artículos, 4 se encuentran en Q1, 2 en Q2, 2 en Q3 y 1 en Q4. Además, se han publicado 3 artículos en revistas no indexadas en el JCR [16,17,24]. Cabe destacar que dos de ellas están indexadas en el *SCImago Journal & Country Rank*. Por último, los resultados obtenidos se han presentado en 7 congresos nacionales e internacionales (con revisión por pares) [3,8,9,12,16,18,19].

Referencias

1. Blum, C., Roli, A.: Metaheuristics in combinatorial optimization: Overview and conceptual comparison. *ACM Comput. Surv.* 35(3), 268 – 308 (2003)
2. Campos, V., Martí, R., Sánchez-Oro, J., Duarte, A.: Grasp with path relinking for the orienteering problem. *J Oper Res Soc* 65(12), 1800 – 1813 (2014)
3. Concha, D., Fernández-López, D., Sánchez-Oro, J., Montemayor, A., Pantrigo, J., Duarte: Búsqueda de vecindad variable secuencial y paralela: una aplicación al problema de la maximización del corte. In: IX Congreso Español sobre Metaheurísticas, Algoritmos Evolutivos y Bioinspirados (MAEB 2013). Madrid (Spain) (2013)
4. Duarte, A., Escudero, L.F., Martí, R., Mladenović, N., Pantrigo, J.J., Sánchez-Oro, J.: Variable neighborhood search for the vertex separation problem. *Computers & Operations Research* 39(12), 3247 – 3255 (2012)

5. Duarte, A., Laguna, M., Martí, R., Sánchez-Oro, J.: Optimization procedures for the bipartite unconstrained 0-1 quadratic programming problem. *Computers & Operations Research* 51(0), 123 – 129 (2014)
6. Duarte, A., Pantrigo, J., Pardo, E., Sánchez-Oro, J.: Parallel variable neighbourhood search strategies for the cutwidth minimization problem. *IMA Journal of Management Mathematics* (2013)
7. Duarte, A., Sánchez-Oro, J., Resende, M., Glover, F., Martí, R.: Grasp with exterior path relinking for differential dispersion minimization. *Information Sciences* 296, 46 – 60 (2015)
8. Fernández-López, D., Sánchez-Oro, J., Duarte, A.: Cálculo de separadores de grafos utilizando búsqueda de vecindad variable reducida. In: IX Congreso Español sobre Metaheurísticas, Algoritmos Evolutivos y Bioinspirados (MAEB 2013). Madrid (Spain) (2013)
9. Fernández-López, D., Sánchez-Oro, J., Duarte, A.: Computing graph separators with variable neighborhood search. In: IX Balkan Conference on Operational Research (BALCOR 2013). Belgrade (Serbia) (2013)
10. Floudas, C., Pardalos, P.M.: *Encyclopedia of Optimization* (2009)
11. Gortázar, F., Duarte, A., Laguna, M., Martí, R.: Black box scatter search for general classes of binary optimization problems. *Computers & Operations Research* 37(11), 1977 – 1986 (2010), metaheuristics for Logistics and Vehicle Routing
12. J., S.O., Duarte, A.: Comparison of different variable neighborhood search strategies for solving the bipartite unconstrained 0-1 quadratic programming problem. In: IX Balkan Conference on Operational Research (BALCOR 2013). Belgrade (Serbia) (2013)
13. Laguna, M., Gortázar, F., Gallego, M., Duarte, A., Martí, R.: A black-box scatter search for optimization problems with integer variables. *Journal of Global Optimization* 58(3), 497 – 516 (2014)
14. Pantrigo, J., Martí, R., Duarte, A., Pardo, E.G.: Scatter search for the cutwidth minimization problem. *Annals of Operations Research* 199(1), 285–304 (2012)
15. Papadimitriou, C.H., Steiglitz, K.: *Combinatorial Optimization: Algorithms and Complexity*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA (1998)
16. Sánchez-Oro, J., Duarte, A.: Grasp with path relinking for the sumcut problem. *International Journal of Combinatorial Optimization Problems and Informatics* 3(1) (2011)
17. Sánchez-Oro, J., Duarte, A.: An experimental comparison of variable neighborhood search variants for the minimization of the vertex-cut in layout problems. *Electronic Notes in Discrete Mathematics* 39(0), 59 – 66 (2012)
18. Sánchez-Oro, J., Duarte, A.: Grasp con path relinking para el problema del sumcut. In: VIII Congreso Español sobre Metaheurísticas, Algoritmos Evolutivos y Bioinspirados (MAEB 2012). Albacete (Spain) (2012)
19. Sánchez-Oro, J., Duarte, A., Laguna, M., Martí, R.: Scatter search aplicado al problema de la minimización del profile de matrices y grafos. In: IX Congreso Español sobre Metaheurísticas, Algoritmos Evolutivos y Bioinspirados (MAEB 2013). Madrid (Spain) (2013)
20. Sánchez-Oro, J., Fernández-López, D., Cabido, R., Montemayor, A., Pantrigo, J.: Radar-based road-traffic monitoring in urban environments. *Digital Signal Processing* 23(1), 364 – 374 (2013)
21. Sánchez-Oro, J., Laguna, M., Duarte, A., Martí, R.: Scatter search for the profile minimization problem. *Networks* 65(1), 10–21 (2015)
22. Sánchez-Oro, J., Mladenović, N., Duarte, A.: General variable neighborhood search for computing graph separators. *Optimization Letters* pp. 1–21 (2014)

23. Sánchez-Oro, J., Pantrigo, J.J., Duarte, A.: Combining intensification and diversification strategies in vns. an application to the vertex separation problem. *Computers & Operations Research* 52, Part B(0), 209 – 219 (2014)
24. Sánchez-Oro, J., Sevaux, M., Rossi, A., Martí, R., Duarte, A.: Solving dynamic memory allocation problems in embedded systems with parallel variable neighborhood search strategies. *Electronic Notes in Discrete Mathematics* 47(0), 85 – 92 (2015)