

Uso de técnicas metaheurísticas avanzadas para resolver problemas de optimización combinatoria

Gabriela Minetti¹, Carolina Salto¹, Hugo Alfonso¹ y Fernando Sanz Troiani²

Laboratorio de Investigación en Sistemas Inteligentes (LISI)

Facultad de Ingeniería - Universidad Nacional de La Pampa

Calle 110 Esq. 9 (6360) General Pico - La Pampa - Rep. Argentina

Te. / Fax: (02302) 422780/422372, Int. 6302

e-mail: ¹{minettig, saltoc, alfonsoh@ing.unlpam.edu.ar}, ²fstnando@gmail.com

Resumen La finalidad de esta línea de investigación es el estudio y resolución de problemas de optimización combinatoria mediante la utilización de métodos aproximados. Particularmente, nuestro trabajo se enfoca en el análisis y desarrollo de algoritmos metaheurísticos basados en trayectoria y en población, así como también híbridos, que permitan resolver eficientemente problemas genéricos como es el caso de QAP y problemas específicos y del mundo real como FAP y TSP. También consideramos la posibilidad de distribuir y/o paralelizar estos métodos dependiendo de la complejidad del problema a resolver.

Palabras claves: Bioinformática, ADN, metaheurísticas, secuenciamiento de un genoma, ensamblado de fragmentos, métodos de búsqueda híbrida y distribuida, optimización combinatoria.

CONTEXTO

El proyecto de investigación en el cual se enmarca esta línea de investigación se denomina "Resolviendo problemas complejos con técnicas metaheurísticas avanzadas". Este proyecto es dirigido por la Dra. Carolina Salto y llevado a cabo en el Laboratorio de Investigación de Sistemas Inteligentes (LISI), de la Facultad de Ingeniería de la Universidad Nacional de La Pampa. Los integrantes de este laboratorio mantienen desde hace varios años una importante vinculación con investigadores de la Universidad Nacional de San Luis (Argentina) y de la Universidad de Málaga (España), con quienes se han realizado varias publicaciones conjuntas.

I. INTRODUCCIÓN

Como es bien conocido, el problema de asignación cuadrática (QAP, del Inglés Quadratic Assignment Problem) es un problema de optimización combinatoria NP-duro que forma parte del núcleo de muchos problemas del mundo real [1]. Diversas aplicaciones en áreas tan diferentes como Investigación Operativa, Optimización Combinatoria y Computación Paralela y Distribuida [2], [3], [4], [5], [6], [7], [8], [9], [10] pueden modelarse por medio de este problema. En particular, el interés de nuestra línea de investigación se centra en los problemas combinatorios, tales como: el problema del viajante de comercio (TSP del Inglés Travelling Salesman Problem), el problema de ensamblado de fragmentos de ADN

(FAP del Inglés Fragment Assembly Problem), diversos problemas de planificación de tareas, entre otros.

En general, la dificultad de este tipo de problemas radica en buscar la solución óptima en un espacio de soluciones de elevada complejidad. En el caso de los métodos de búsqueda exactos, esta dificultad se traduce en tiempos de cómputos inviables para hallar una solución óptima. En cambio, los métodos aproximados, especialmente las metaheurísticas, reducen considerablemente la complejidad temporal para obtener soluciones óptimas o quasi-óptimas. En consecuencia, esto implica una reducción significativa de los costos de cualquier organización a la hora de resolver esta clase de problemas.

Las metaheurísticas obtienen soluciones que cumplen con la calidad requerida y los tiempos de demora impuestos en el campo industrial, además permiten estudiar clases genéricas de problemas en lugar de instancias de problemas particulares. En general, las técnicas que obtienen mejores resultados (en precisión y esfuerzo) al resolver problemas complejos y del mundo real usan metaheurísticas. Sus campos de aplicación van desde la optimización combinatoria, bioinformática, telecomunicaciones a la economía, software ingeniería, etc, que necesitan soluciones rápidas con alta calidad [11].

Existen diferentes formas de clasificar y describir las técnicas metaheurísticas [12], [13], [14]. Dependiendo de las características que se seleccionen se pueden obtener diferentes taxonomías: basadas en la naturaleza y no basadas en la naturaleza, con memoria o sin ella, con una o varias estructuras de vecindario, etc. Una de las clasificaciones más populares las divide en *metaheurísticas basadas en trayectoria* y *basadas en población*. Las primeras realizan búsquedas orientadas a la explotación del espacio de soluciones al manipular en cada paso un único elemento de dicho espacio. En tanto que, las segundas son métodos orientados a la exploración del espacio de búsqueda, ya que trabajan sobre un conjunto de elementos (población). Esta taxonomía se muestra de forma gráfica en la figura 1, que además incluye las principales metaheurísticas.

En los últimos años, el interés en combinar metaheurísticas (*metaheurísticas híbridas*) ha aumentado considerablemente en el área de optimización [15]. Combinaciones de algorit-

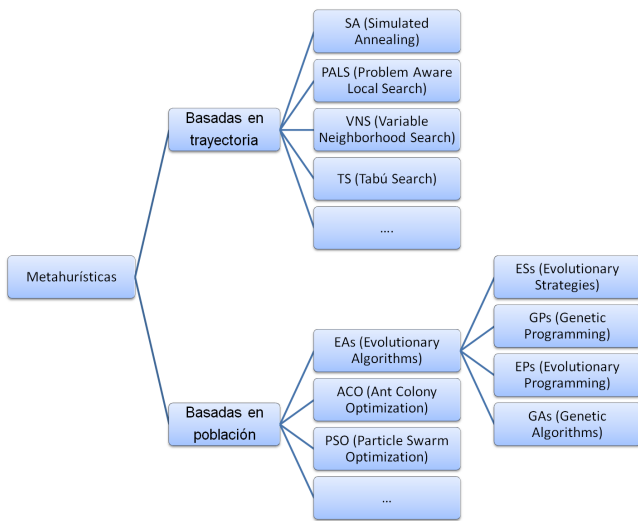


Fig. 1. Clasificación de las metaheurísticas.

mos tales como metaheurísticas basadas en población y/o en trayectoria, programación matemática, programación con restricciones y técnicas de aprendizaje automático (machine learning) proveen algoritmos de búsqueda muy poderosos.

Los objetivos de esta línea de investigación son: analizar, diseñar y desarrollar algoritmos metaheurísticos basados en trayectoria y en población, así como también híbridos, que permitan resolver eficientemente problemas genéricos como es el caso de QAP y problemas específicos y del mundo real como FAP y TSP.

II. DESARROLLO

En esta sección describimos los desarrollos que se llevan a cabo en esta línea de investigación, pero primero introducimos una breve descripción de cada uno de los problemas a resolver.

El Problema de Asignación Cuadrática es un problema clásico de optimización combinatorio, en el cual se encuentra un vasto número de problemas de diseño y de distribución de recursos en diferentes campos, donde la decisión a tomar es una asignación de elementos de un conjunto en otro. El QAP es considerado como un problema complejo y dificultoso de resolver y puede establecerse como un conjunto de n elementos distintos que deben ser localizados (asignados) en n localidades distintas minimizando el costo. En cada localidad sólo puede haber un elemento y todos los elementos deben ser asignados a una localidad. Por cada par de localidades el costo es calculado como el producto de la distancia entre las localidades y el flujo asociado a los elementos en las localidades. El costo total es la suma de todos los costos asociados a cada par de localidades.

En tanto que, FAP es un problema resuelto en las primeras fases del proyecto del genoma y por lo tanto muy importante, ya que los demás pasos dependen de su precisión. El proceso de ensamblado de fragmentos consiste en: una primera fase de superposición (calcula el *puntaje de solapamiento* entre los fragmentos), una segunda de distribución (encuentra el

orden de los fragmentos basado en el puntaje de similitud computado) y una tercera de consenso (deriva la secuencia de ADN a partir de la distribución anterior). Una resolución óptima del problema se produce cuando el algoritmo es capaz de ensamblar un determinado conjunto de fragmentos en un solo *contig*. Un *contig* es una secuencia en la que el solapamiento entre los fragmentos adyacentes es mayor o igual a un umbral predefinido (parámetro de corte denominado *cutoff*).

El problema del viajante de comercio, es uno de los problemas de optimización combinatoria NP-duros más ampliamente estudiado. Su declaración es engañosamente simple: un viajante busca el camino más corto para pasar por m ciudades. En otras palabras, una persona debe visitar un conjunto de m ciudades, comenzando en una ciudad determinada y finalizando en la misma ciudad; luego de haber visitado todas ellas sólo una vez. Esto significa que nunca regresa a una ciudad ya visitada, excepto la primera.

Desde el punto de vista de la optimización combinatoria, la construcción de un consenso (en FAP) es similar a la de un recorrido en un problema del viajante de comercio. Esto es porque cada fragmento tiene una ubicación específica en la formación de una secuencia en la etapa de consenso. Aunque los puntos terminales de un recorrido de TSP sean irrelevantes ya que su solución es un recorrido circular de ciudades, en el caso de FAP estos puntos son importantes ya que ellas representan los extremos opuestos de la secuencia original de ADN. En TSP el ordenamiento de las ciudades es la solución final al problema. En cambio para FAP, el ordenamiento de fragmentos es sólo un resultado intermedio que será utilizado en la fase de consenso. Por otra parte, estos dos problemas son formulados como casos especiales de QAP.

Por un lado, nuestro trabajo consiste en profundizar el estudio sobre la robustez de las metaheurísticas poblacionales basados en colonia de hormigas (*Ant Colony Optimization* -ACO-) [16] para resolver TSP. Con el propósito de estudiar la robustez de una metaheurística se analizan las diferencias entre las soluciones encontradas para las instancias sin y con ruido. Si no se detectan diferencias (estadísticamente significativas), la metaheurística muestra un comportamiento neutro (insensible, indistinto) a pequeñas variaciones en los datos de entrada. Consecuentemente, esta metaheurística se considera robusta para resolver instancias ruidosas. De esta forma, si ACO resultara una metaheurística robusta para TSP podríamos adaptarla fácilmente para resolver las instancias con ruido de FAP.

Por otra parte analizamos el comportamiento de los ensambladores metaheurísticos que hasta el momento han mostrado un mejor desempeño: ISA (*Inversion Simulated Annealing*), PALS (*Problem Aware Local Search*) y GAG₅₀ (*Genetic Algorithm with a Greedy seeding strategy*) [17], [18], [19]. Los ensambladores ISA, PALS y GAG₅₀ han logrado una mejor calidad de los resultados, ya sea medida por medio del *fitness* o por el número final de contigs. Además, han sido capaces de resolver eficazmente un conjunto completo de instancias

con un esfuerzo computacional mínimo [17], [19], [20], [21]. Teniendo en cuenta las virtudes de estos ensambladores desarrollamos una nueva metaheurística híbrida para resolver instancias de mayor tamaño de FAP, denominada SAX. El objetivo perseguido con SAX es combinar la potencialidad de ISA como ensamblador de fragmentos con operadores de recombinación genéticos como intensificadores de la búsqueda.

Finalmente, dado que SAX y PALS demostraron resolver eficientemente el problema de ensamblado de fragmentos [22], [18] analizaremos el comportamiento de estas dos metaheurísticas en la resolución del problema genérico QAP.

III. RESULTADOS OBTENIDOS/ESPERADOS

En esta sección presentamos los resultados obtenidos de nuestra investigación en el transcurso del año 2012 y los esperados en el 2013.

Con respecto al análisis de robustez de ACO, frente a la presencia de ruido en instancias de TSP, esperamos encontrar evidencia empírica que demuestre fehacientemente dicha característica. Estos resultados nos permitirán una mejor adaptación de esta técnica metaheurística a la resolución de instancias con o sin ruido de FAP.

Por otra parte, desarrollamos un nuevo ensamblador, denominado SAX, para resolver FAP [22]. Este ensamblador es la combinación de una heurística basada en trayectoria (ISA) con un operador genético de cruce basado en el orden (OX, *Order Crossover*). El componente ISA le permite a SAX utilizar el procedimiento por inversión para generar dos nuevas soluciones y así extender la exploración hacia regiones de búsqueda prometedoras, además de, utilizar el criterio de Boltzman para escapar de óptimos locales. El componente OX incrementa la explotación de las regiones prometedoras en el espacio de búsqueda. De esta forma, SAX encuentra distribuciones óptimas con altos puntajes de superposición para todas las instancias de FAP.

Finalmente esperamos encontrar resultados que demuestren empíricamente que SAX y PALS son aplicables al grupo de problemas derivados de QAP. Es importante destacar que, tanto en esta investigación como en las dos anteriores, las conclusiones obtenidas del análisis de resultados son verificadas rigurosamente a través de los correspondientes tests estadísticos.

IV. FORMACIÓN DE RECURSOS HUMANOS

Durante el año 2012, integrantes del proyecto han realizado diversos cursos de postgrado directamente relacionados con la temática del proyecto, con el objetivo de llevar a cabo carreras de postgrado en un futuro cercano.

En tanto que, en el LISI se trabaja con alumnos avanzados en la carrera Ingeniería en Sistemas en temas relacionados a la resolución de problemas de optimización usando técnicas inteligentes, con el objeto de guiarlos en el desarrollo de sus tesis de grado y, también, de formar futuros investigadores.

REFERENCES

- [1] M. Garey and D. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, 1979.
- [2] R. S. Bhaba, E. W. Wilbert, and L. H. Gary, "Locating sets of identical machines in a linear layout," *Ann. Oper. Res.*, vol. 77, pp. 183–207, 1998.
- [3] S. P. Coy, B. L. Golden, G. C. Runger, and E. A. Wasil, "See the forest before the trees: fine-tuned learning and its application to the traveling salesman problem," *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, pp. 454–464, 1998.
- [4] D. Chiang and L. C. Potter, "Minimax non-redundant channel coding for vector quantization," in *Proceedings of the 1993 IEEE international conference on Acoustics, speech, and signal processing: image and multidimensional signal processing - Volume V*, ser. ICASSP'93. Washington, DC, USA: IEEE Computer Society, 1993, pp. 617–620. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1947148.1947321>
- [5] D. Gamboa, C. Rego, and F. Glover, "Implementation analysis of efficient heuristic algorithms for the traveling salesman problem," *Comput. Oper. Res.*, vol. 33, no. 4, pp. 1154–1172, Apr. 2006. [Online]. Available: <http://dx.doi.org/10.1016/j.cor.2005.06.014>
- [6] A. Mason and M. Rönnqvist, "Solution methods for the balancing of jet turbines," *Computers & Operations Research*, vol. 24, pp. 153–167, 1998.
- [7] J. W. Pepper, B. L. Golden, and E. A. Wasil, "Solving the traveling salesman problem with annealing-based heuristics: a computational study," *Trans. Sys. Man Cyber. Part A*, vol. 32, no. 1, pp. 72–77, Jan. 2002. [Online]. Available: <http://dx.doi.org/10.1109/3468.995530>
- [8] G. G. Polak, "On a special case of the quadratic assignment problem with an application to storage-and-retrieval devices," *Annals of Operations Research*, vol. 138, pp. 223–233, 2005.
- [9] R. C., "Relaxed tours and path ejections for the traveling salesman problem," *European Journal of Operational Research*, vol. 106, no. 2, pp. 522–538, 1998.
- [10] Q. Zhang, J. Sun, and E. Tsang, "An evolutionary algorithm with guided mutation for the maximum clique problem," *IEEE Transactions on Evolutionary Computation*, vol. 9, no. 2, pp. 192–200, April 2005.
- [11] E. Alba, G. Luque, and S. Nesmachnow, "Parallel metaheuristics: recent advances and new trends," *International Transactions in Operational Research*, vol. 20, no. 1, pp. 1–48, 2013. [Online]. Available: <http://dx.doi.org/10.1111/j.1475-3995.2012.00862.x>
- [12] E. Cantú-Paz, "Migration, selection pressure, and superlinear speedups," in *Efficient and Accurate Parallel Genetic Algorithms*. Kluwer Academic Publishers, 2000, pp. 97–120.
- [13] T. G. Crainic and M. Toulouse, "Parallel strategies for metaheuristics," in *Handbook of Metaheuristics*. Kluwer Academic Publishers, 2003, pp. 475–513.
- [14] C. Blum and A. Roli, "Metaheuristics in combinatorial optimization: Overview and conceptual comparison," *ACM Comput. Surv.*, vol. 35, pp. 268–308, September 2003.
- [15] E.-G. Talbi, "A taxonomy of hybrid metaheuristics," *Journal of Heuristics*, vol. 8, pp. 541–564, September 2002. [Online]. Available: <http://portal.acm.org/citation.cfm?id=594957.595102>
- [16] M. Dorigo, "Optimization, learning and natural algorithms," Ph.D. dissertation, Politecnico di Milano, Italy, 1992.
- [17] G. Minetti, G. Luque, E. Alba, and G. Leguizamón, "A new Hybrid SA for Solving the DNA Fragment Assembly Problem," in *XXVIII International Conference of the Chilean Computing Science Society (SCCC)*, November 2009, pp. 109–116.
- [18] E. Alba and G. Luque, "A new local search algorithm for the dna fragment assembly problem," in *Evolutionary Computation in Combinatorial Optimization, EvoCOP'07*, ser. Lecture Notes in Computer Science. Valencia, Spain: Springer, 2007, vol. 4446, pp. 1–12.
- [19] G. Minetti, E. Alba, and G. Luque, "Seeding strategies and recombination operators for solving the DNA fragment assembly problem," *Information Processing Letters*, vol. 108, no. 3, pp. 94–100, October 2008.
- [20] G. Minetti and E. Alba, "Metaheuristic assemblers of DNA strands: Noiseless and noisy cases," in *2010 IEEE Congress on Evolutionary Computation (CEC)*, July, pp. 1–8.
- [21] G. Minetti, G. Leguizamón, and E. Alba, "Assembling DNA Sequences Containing Noisy Information With Metaheuristic Algorithms," *Journal of Information Sciences, Elsevier (in evaluation)*, 2011.

- [22] G. F. Minetti, M. G. Leguizamón, and E. A. Torres, "SAX: a new and efficient assembler for solving DNA Fragment Assembly Problem," in *41JAIIO Jornadas Argentinas de Informática*, S. A. de Informática (SADIO), Ed. SADIO, 2012, pp. 177–188.