

Metaheurísticas aplicadas a problemas de optimización

Hugo Alfonso¹, Carolina Salto¹, Gabriela Minetti¹, Natalia Stark¹
Carlos Bermúdez², Alina Orellana³, Fernando Sanz Troiani⁴

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{alfonsoh, saltoc, minettig, nstark@ing.unlpam.edu.ar },
²bermudezc@yahoo.com, ³orellanaalina@gmail.com, ⁴fstnando@hotmail.com

RESUMEN

El objetivo principal de esta línea de investigación es el diseño y desarrollo de algoritmos heurísticos y meta-heurísticos que resuelvan problemas de optimización. En particular se abordan problemas de corte y empaquetado, de ruteo de vehículos y de ensamblado de fragmentos de ADN. Actualmente dos de las ramas con más éxito para diseñar metaheurísticas eficientes, y dar solución a estos problemas, son la hibridación y el paralelismo. El trabajo está orientado a aplicar metaheurísticas secuenciales y paralelas a los problemas propuestos, a analizar los resultados para comprender el comportamiento de estos algoritmos y a proponer nuevos métodos para resolver los problemas de una manera más eficaz y eficiente.

Palabras clave: *metaheurísticas, computación paralela, optimización combinatoria, métodos de búsqueda híbrida.*

CONTEXTO

Las líneas de investigación descritas en esta presentación se enmarcan en el Proyecto de Investigación “*Sistemas Metaheurísticos para Resolver Problemas de Optimización*”, llevado a cabo en el Laboratorio de Investigación de Sistemas Inteligentes (LISI), de la Facultad de Ingeniería de la Universidad Nacional de La Pampa.

1. INTRODUCCION

Tradicionalmente, uno de los aspectos más importante de la investigación en informática es el diseño de algoritmos cada vez más eficientes para resolver problemas complejos. Fundamentalmente, el objetivo perseguido en este campo es el desarrollo de nuevos métodos capaces de resolver dichos problemas complejos con el menor esfuerzo computacional posible, mejorando así a los algoritmos existentes. Esto no sólo permite

afrontar problemas actuales de forma más eficiente, sino también tareas que en el pasado parecían imposibles de solucionar debido a su alto costo computacional. En este contexto, la actividad investigadora está creciendo de forma evidente en estos días, tanto en el planteo de algoritmos exactos como en heurísticos *ad hoc* y metaheurísticos para resolver problemas complejos de optimización. La razón es que continuamente se están afrontando nuevos problemas de ingeniería, mientras que, al mismo tiempo, cada vez se dispone de mejores recursos computacionales, como nuevos tipos de computadoras, redes, y entornos como Internet.

Los algoritmos exactos garantizan encontrar para cada instancia del problema de tamaño finito una solución óptima en tiempo limitado (ver [16 y 17]), pero tienen el grave inconveniente de que en problemas reales (que suelen ser NP-duros [8] en la mayoría de los casos) su tiempo de ejecución crece de forma exponencial con el tamaño del problema. Por ello, los algoritmos heurísticos *ad hoc* han recibido mucha atención, ya que sacrifican la garantía de alcanzar soluciones óptimas por la de encontrar buenas soluciones en un tiempo significativamente reducido. Otro inconveniente de los heurísticos *ad hoc* es que no son fáciles de definir en determinados problemas. Las metaheurísticas [3,6,9] ofrecen un equilibrio adecuado entre ambos extremos: son métodos genéricos que ofrecen soluciones de buena calidad (el óptimo global en muchos casos) en un tiempo moderado.

Las líneas de investigación desarrolladas en nuestro proyecto se han centrado en el estudio de metaheurísticas basados en trayectoria y basadas en población. Las primeras realizan un proceso de búsqueda caracterizado por determinar una trayectoria en el espacio de soluciones. Por lo tanto, una solución sucesora puede o no pertenecer al vecindario de la actual. Los métodos basados en población operan en cada iteración del algoritmo

con un conjunto de soluciones, denominada población. Así ellos, en forma natural, exploran el espacio de búsqueda desde varias regiones a la vez. El resultado final depende de la forma en que se manipula la población.

Dos de las ramas más exitosas para diseñar algoritmos eficientes en la actualidad son la hibridación y el paralelismo [2]. La hibridación permite incorporar información del problema en el algoritmo de resolución para trabajar en contacto con sus características diferenciadoras; esto tiene también relación con la posibilidad de involucrar a varios algoritmos distintos en el proceso de búsqueda de una solución de manera más eficiente [7]. Utilizar algoritmos paralelos es una forma de aliviar los problemas vinculados a tiempos de ejecución intensivos y requerimientos de memoria importantes para resolver instancias complejas de interés actual. Varias estrategias de paralelización se pueden aplicar a las metaheurísticas. Entre ellas, existen modelos paralelos donde varios hilos de búsqueda simultáneamente exploran el espacio de soluciones. Si cada uno usa diferentes procedimientos de búsqueda, obtenemos una metaheurística heterogénea paralela. La utilización de múltiples hilos de búsqueda usando diferentes estrategias y valores de los parámetros permiten una mayor diversidad y una exploración más profunda del espacio de búsqueda, lo que podría llevar a soluciones más precisas.

El objetivo de las líneas de investigación, que se vienen desarrollando desde hace más de una década, consiste en aplicar técnicas metaheurísticas a problemas de optimización tanto presentes en el mundo real como aquellos que revisten un carácter teórico pero con aplicaciones en varios campos, analizando distintas posibilidades para sacar el máximo partido a dichas técnicas y ofrecer así soluciones de gran calidad con recursos computacionales al alcance de cualquier institución. Los problemas abordados son concretamente:

- *problema de ruteo de vehículos* [11]: consiste en entregar artículos, utilizando las rutas de costo mínimo, a un determinado conjunto de clientes (de los que conocemos sus demandas), comenzando y acabando en el almacén. Hay un gran número de extensiones del VRP canónico. Una extensión básica se conoce como VRP con Capacidad limitada (CVRP), en el que los vehículos tienen capacidades fijas.
- *problema de ensamblado de fragmentos de ADN*. Para enunciar este problema, es necesario definir previamente el proceso de secuenciamiento [18]: 1) El ADN es dividido aleatoriamente en millones de fragmentos, 2) dichos fragmentos son leídos por una máquina de secuenciamiento de ADN y 3) un ensamblador une los fragmentos leídos que se superponen, reconstruyendo la secuencia original. Esta es una técnica general denominada *shotgun sequencing* e introducida por Sanger et al. [23]. El ensamblado de fragmentos de ADN se divide en tres fases diferentes: fase de superposición (encuentra los fragmentos superpuestos), fase de distribución (encuentra el orden de los fragmentos basado en el puntaje de similitud computado) y por último, la fase de consenso (deriva la secuencia de ADN a partir de la distribución anterior).
- *problema de corte y empaquetado*: consiste en el corte de materias primas para obtener un conjunto de elementos minimizando el desperdicio de material generado o en el empaquetado de un conjunto de artículos en el menor número de contenedores. En particular, nos hemos enfocado en una de las variantes más conocidas del problema como lo es el *strip packing*, en el cual la característica más importante es que la plancha de material, donde se deben asignar las distintas piezas rectangulares, tiene una de las dimensiones libres (por lo general la altura) [10].
- *NK-landscapes* es una función de fitness $f: \{0,1\}^N \rightarrow \mathfrak{R}$ definida sobre cadenas binarias [12], donde N es el largo de la cadena y K es la cantidad de bits en dicha cadena que interactúan epistáticamente con un bit dado. El modelo NK es útil como un modelo general tanto para investigar las propiedades estructurales del espacio de búsqueda como para evaluar el comportamiento de un algoritmo evolutivo. También ha servido como punto de referencia para entender las propiedades de algunos objetos del campo de la biología (secuencia de aminoácidos, secuencias RNA, evolución molecular). La evolución de organizaciones en ambientes empresariales se puede modelar basándose en el modelo NK.
- *Problema de corte máximo* (maxCut) consiste en particionar un conjunto de vértices de un

grafo ponderado en dos conjuntos disjuntos, tal que la suma de los pesos de los vértices con un extremo en cada subconjunto es maximizada. Además de su importancia teórica, este problema tiene muchas aplicaciones en varios campos tal como diseño de redes, diseño de circuitos VLSI, clustering de datos, etc.

2. LINEAS DE INVESTIGACION y DESARROLLO

Esta sección se divide atendiendo a las grandes líneas de investigación y desarrollo que se desarrollan en el proyecto:

- Estudio, diseño y aplicación de metaheurísticas secuenciales, híbridas y descentralizadas que permitan mejorar la eficacia y la eficiencia en el manejo y el uso de la información generada por la biología molecular; particularmente, para estudiar la información estructural de una secuencia desconocida de ADN y así resolver el problema de ensamblado de fragmentos.
 - Recopilación y estudio de la bibliografía existente relacionada con el tema de estudio.
 - Identificación de las características específicas del problema de ensamblado de fragmentos.
 - Diseño y desarrollo de nuevos algoritmos metaheurísticos, los cuales pueden ser híbridos y/o descentralizados.
 - Análisis y evaluación de los algoritmos obtenidos para guiar su perfeccionamiento y posibles nuevas propuestas.
 - Obtención, análisis y difusión de los resultados.
- Algoritmos genéticos celulares para resolver problemas de ruteo, tanto estáticos como dinámicos.
 - Estudio del estado del arte.
 - Desarrollo de un operador de recombinación que incorpora información del problema.
 - Comparación con operadores tradicionales de recombinación propuestos para permutaciones.
 - Estudio de las variantes dinámicas del problema.
 - Análisis de casos de pruebas de problemas dinámicos.
 - Adaptación del algoritmo celular para resolver el problema de ruteo dinámico seleccionado.
 - Desarrollo y aplicación de diferentes métricas que permitan estudiar el comportamiento del algoritmo celular.
- Algoritmos evolutivos paralelos heterogéneos aplicados a la resolución de problemas NK-landscapes y Max-Cut.
 - Estudio del estado del arte

- Propuestas de versiones secuenciales con adaptación de los operadores en función del progreso del proceso de búsqueda.
- Desarrollo de algoritmos heterogéneos basados en la configuración de parámetros (subpoblaciones con diferentes valores de probabilidades de los operadores).
- Desarrollo de algoritmos heterogéneos basados en operadores (subpoblaciones utilizando distintos operadores de recombinación).
- Profundización en el estudio de las distintas variantes en algoritmos de colonia de hormigas.
 - Análisis de las características diferenciadoras de las distintas variantes algorítmicas existentes en la literatura.
 - Desarrollo e implementación de cada una de ellas.
 - Determinación de los casos de prueba para testear las propuestas algorítmicas.
 - Análisis comparativo para extraer conclusiones.
- Estudio, diseño y aplicación de la metaheurística optimización por cúmulo de partículas.
 - Análisis de las distintas propuestas bibliográficas que aplican esta metaheurística para resolver problemas con representación binaria.
 - Análisis, diseño y desarrollo de nuevas modificaciones de este algoritmo para incrementar la eficiencia en la optimización de problemas con representación binaria.
 - Aplicación de estos resultados en la resolución del problema de diseño de redes de antenas de radio.
 - Análisis y evaluación de los resultados obtenidos para mejorar el desempeño de este algoritmo en la resolución del problema antes mencionado.
 - Evaluación, síntesis y difusión de los resultados finales.

3. RESULTADOS OBTENIDOS/ESPERADOS

Durante el último año el grupo ha profundizado en el estudio e implementación de diversas metaheurísticas, cuyo rendimiento se ha evaluado sobre distintos problemas (ver Sección 2).

Hemos analizado, adaptado y comparado distintas metaheurísticas (algoritmos genéticos y simulated annealing) y métodos de búsqueda local (Problem Aware Local Search) [14] para resolver el problema de ensamblado de fragmentos de ADN. Se utilizaron instancias tradicionales del problema y también nuevos benchmarks de mayor complejidad; estos últimos fueron obtenidos por nuestro grupo de trabajo. También hemos

estudiado cómo los algoritmos anteriores afectan la composición estructural de la solución de instancias del problema con y sin ruido [15]. En función a estos resultados se espera desarrollar un algoritmo metaheurístico que incorpore las ventajas de los algoritmos antes mencionados, además de ser resistente a datos de entrada ruidosos.

Hemos realizado un estudio de distintas alternativas de GA para seleccionar una configuración adecuada al problema de strip packing. Los resultados muestran que un algoritmo genético usando operadores específicos al problema ha resultado ser muy eficiente para resolver el problema comparado con GA aplicando operadores tradicionales de permutaciones, no sólo por la bondad de sus resultados sino también por el menor número de evaluaciones que necesita [19,20]. Las características de los modelos distribuidos propuestos, tanto de GA como de ACS, han mostrado ser técnicas rápidas que producen buenos patrones de empaquetado. Todos estos resultados han sido recopilados en una tesis doctoral presentada en 2009 [21].

El nuevo operador de recombinación propuesto para el CVRP ha mostrado dar buenos resultados para mejorar la calidad de soluciones frente a operadores tradicionales de recombinación para permutaciones utilizados en la literatura. Este operador se incorporó a un algoritmo genético celular que recientemente obtuvo los mejores resultados reportados en la literatura [1]. Podemos considerar que el salto cualitativo logrado en la resolución de este problema con la utilización de un algoritmo celular fue importante respecto de un algoritmo genético [4]. El algoritmo celular propuesto usando operadores que incorporan conocimiento específico del problema fue mejorado con un operador de búsqueda local. Comparando el comportamiento de este algoritmo con uno presente en la literatura [1], nuestra propuesta obtiene buena calidad de soluciones con altos niveles de éxito [5]. Con el conocimiento desarrollado en ésta área, planeamos incursionar en problemas de ruteo dinámico, en los cuales los clientes pueden presentar sus órdenes en cualquier momento de la búsqueda. Para ello utilizaremos un algoritmo celular que tan buenos resultados ha dado en la versión estática de VRP.

Las características de los algoritmos paralelos heterogéneos propuestos basados en operadores desarrollado para resolver el modelo NK han mostrado ser técnicas rápidas que obtienen buenos

resultados, lo cual representa un avance prometedor en esta área. Este algoritmo ha obtenido un rendimiento numérico más alto con mejores niveles de exactitud con respecto a un algoritmo paralelo homogéneo [22]. Estos resultados indican que la interacción de diferentes hilos de búsqueda con distintos parámetros da lugar a nuevas técnicas de búsqueda prometedoras. Nos proponemos analizar otro nivel de heterogeneidad (para ello utilizaremos el problema de corte máximo), incluida la adaptación dinámica de las probabilidades de los operadores de acuerdo con el progreso de la búsqueda evolutiva. Por otra parte, estamos interesados en analizar mejor el complejo comportamiento de los algoritmos paralelos heterogéneos para entender la relación interna de la combinación de diferentes operadores genéticos durante el proceso de búsqueda.

Con respecto al estudio de algoritmos de colonia de hormigas se espera conocer dentro de qué rangos los parámetros de estos algoritmos deben estar configurados para resolver óptimamente distintas clases de problemas. También se pretende difundir los resultados anteriores a nivel nacional.

Hemos propuesto una nueva versión de la metaheurística de optimización por cúmulo de partículas; la cual incorpora características específicas de los algoritmos genéticos en general y de CHC en particular [13]. Durante el 2010 se está trabajando en estrategias para mejorar la creación de soluciones iniciales correspondientes al problema de diseño de redes de antenas de radio.

4. FORMACION DE RECURSOS HUMANOS

Actualmente se está desarrollando una tesis de Doctorado en Ciencias de la Computación dentro del ámbito de la Universidad Nacional de San Luis, estando prevista su conclusión durante el presente año. Además durante 2009 se ha concluido la escritura de una tesis doctoral y se ha realizado la correspondiente presentación y defensa para obtener el título de Doctor en Ciencias de la Computación (UNSL).

Por otra parte, a lo largo del año 2009 uno de los becarios del proyecto ha presentado su tesis para alcanzar el título de Ingeniero en Sistemas, a partir de las actividades concretas que desarrollaron en el LISI. Además, actualmente otro de los becarios está desarrollando su trabajo de tesis para el grado de Ingeniero en Sistemas, a partir de las actividades que viene desarrollando en el proyecto.

5. BIBLIOGRAFIA

- [1] E. Alba and B. Dorronsoro. A hybrid cellular genetic algorithm for the capacitated vehicle routing problem. In Ajith Abraham, Crina Grosan, and Witold Pedrycz, editors, *Engineering Evolutionary Intelligent Systems*, volume 82 of *Studies in Computational Intelligence*, pages 379–422. Springer, 2008.
- [2] S.G. AKL. *Diseño y Análisis de Algoritmos Paralelos*. RA-Ma, 1992.
- [3] T. Bäck. *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming*. Oxford University Press, New York, 1996.
- [4] C. Bermudez, C. Salto, A. Alfonso, Algoritmos Celulares con Operadores Específicos para Resolver un Problema de Ruteo de Vehículos. En los anales del XIV Congreso Argentino de Ciencias de la Computación (CACIC 2009), 2009.
- [5] C. Bermudez, P. Graglia, N. Stark, C. Salto, H. Alfonso, Comparison of Recombination Operators in Panmictic and Cellular GAs to Solve a Vehicle Routing Problem, en *Revista Iberoamericana de Inteligencia Artificial*, vol. 46, pág. 34-44, marzo 2010.
- [6] C. Blum and A. Roli. Metaheuristics in combinatorial optimization: Overview and conceptual comparison. *ACM Computing Surveys*, 35(3):268–308, 2003.
- [7] L. Davis. *Handbook of Genetic Algorithms*. Van Nostrand Reinhold, 1991.
- [8] M. Garey y D. Johnson. *Computers and Intractability: a Guide to the Theory of NP-completeness*. Freeman, San Francisco, California, 1979.
- [9] D. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, Reading, MA, 1989.
- [10] E. Hopper. Two-dimensional packing utilising evolutionary algorithms and other metaheuristic methods. PhD thesis, University of Wales, Cardiff, U.K., 2000.
- [11] G. B. Dantzig and J. H. Ramser. The truck dispatching problem. *Management Science*, 6(1):80–91, 1959.
- [12] S. Kauffman. *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, 1993.
- [13] A. Orellana, G. Minetti, A Modified Binary-PSO for Continuous Optimization, en los Proceedings del XV Congreso Argentino de Ciencias de la Computación (CACIC 2009), 2009.
- [14] G. Minetti, G. Luque, E. Alba, G. Leguizamón, A new Hybrid SA for Solving the DNA Fragment Assembly Problem, en Proceedings del XXVIII Internacional Conference of the Chilean Computing Science Society (SCCC). 2009.
- [15] G. Minetti, E. Alba, Metaheuristic assembler of the DNA strands: noiseless and noisy cases. In Proceedings Congress on Evolutionary Computation, 2010. en prensa.
- [16] G. Nemhauser y L. Wolsey. *Integer and Combinatorial Optimization*. John Wiley, New York, 1998.
- [17] C.H. Papadimitriou y I.C. Steiglitz. *Combinatorial Optimization: Algorithms and Complexity*. Prentice-Hall, Englewood Cliffs edition, 1982.
- [18] M. Pop, S.L. Salzberg, and M. Shumway. Genome sequence assembly: Algorithms and issues. *Computer*, 35(7):47–54, 2002.
- [19] C. Salto, J.M. Molina y E. Alba. Greedy Seeding and Problem Specific Operators for GAs Solving Strip Packing Problems. In "Optimization Techniques for Solving Complex Problems", E. Alba, C. Blum, P. Isasi, C. León, and J.A. Gómez (Editors), A Wiley Interscience Publication, JOHN WILEY & SONS, (in Press). ISBN: 978-0-470-29332-4. pág 385-405, 2009
- [20] C. Salto, G. Leguizamón, E. Alba, y J.M. Molina. Evolutionary and ACO Based Approaches for Two-dimensional Strip Packing Problem. In "Natural Intelligence for Scheduling, Planning, and Packing Problems", Springer-Verlag in the series "Studies in Computational Intelligence". Raymond Chiong (editor). In press. ISBN: 978-3-642-04038-2, pág. 245-266, 2009
- [21] C. Salto, *Metaheurísticas híbridas-paralelas para problemas industriales de corte, empaquetado y otros relacionados*". Tesis de doctoral, Universidad Nacional de San Luis, 2009.
- [22] C. Salto, E. Alba, Heterogeneous Parallel Algorithms to solve Epistatic Problems, Proceedings of 13th International Workshop on Nature Inspired Distributed Computing (NIDISC'10), En edición, 2010.
- [23] F. Sanger, A.R. Coulson, G.F. Hong, D.F. Hill, and G.B. Petersen. Nucleotide Sequence of Bacteriophage Lambda DNA. *Journal of Molecular Biology*, 162(4):729–773, 1982.