

Sintonización automática de aplicaciones paralelo/distribuidas basadas en algoritmos evolutivos y evolución diferencial

Caymes-Scutari Paola^{1,2}, Bianchini Germán¹, Tardivo María Laura^{1,2,3},
Méndez-Garabetti Miguel^{1,2}

¹Laboratorio de Investigación en Cómputo Paralelo/Distribuido
Departamento de Ingeniería en Sistemas de Información
Facultad Regional Mendoza/Universidad Tecnológica Nacional
Rodríguez 273 (M5502AJE) Mendoza, +54 261 5244579

²Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

³Departamento de Computación, Universidad Nacional de Río Cuarto, Río Cuarto,
Córdoba, Argentina

pcaymesscutari@frm.utn.edu.ar, gbianchini@frm.utn.edu.ar, mmendez@mendoza-conicet.gob.ar,
lauratardivo@dc.exa.unrc.edu.ar

RESUMEN

El procesamiento paralelo se ha convertido en un recurso muy utilizado para el procesamiento de datos y la resolución de problemas en diversos campos de la ciencia y la ingeniería. Sin embargo, el paradigma paralelo conlleva una complejidad considerable dado que involucra un gran número de aspectos no funcionales (como granularidad, técnica de descomposición y/o asignación, grado de concurrencia, etc.) que pueden impactar negativamente en el rendimiento de las aplicaciones. Mitigar dicho efecto negativo requiere de un cierto dominio y experiencia en su manejo, lo cual suele resultar prohibitivo para el usuario no experto en informática y en paralelismo en particular. Es por ello que tanto el proceso de desarrollo de aplicaciones paralelas como el proceso de sintonización de las mismas constituyen importantes desafíos a la hora de abordar un problema de manera paralela y eficiente. El Proceso de Sintonización

comporta una serie de etapas sucesivas mediante las cuales las aplicaciones son instrumentadas, monitorizadas, analizadas y sintonizadas (o ajustadas) de acuerdo a las características del problema que presentan o al entorno de ejecución, con el fin de mejorar su rendimiento y hacer un mejor aprovechamiento de los recursos. Sin lugar a dudas, el proceso de sintonización requiere un grado de conocimiento de la aplicación, el paradigma paralelo y los problemas de sintonización que usualmente lo vuelve una tarea muy difícil y restrictiva para lograr un rendimiento adecuado, sobre todo cuando de usuarios no expertos se trata. El presente proyecto propone abordar la sintonización automática de aplicaciones paralelas basadas en Evolución Diferencial.

Palabras clave: Evolución Diferencial, Algoritmos Evolutivos, Sintonización Automática, HPC, Problemas de Rendimiento, Problem Solver.

CONTEXTO

Este proyecto se enmarca en una línea de investigación que se viene desarrollando en los últimos años en el ámbito del LICPaD (Laboratorio de Investigación en Cómputo Paralelo/Distribuido) en la UTN-FRM. La misma se centra en el desarrollo de un entorno denominado EDDSAP (Entorno de Especificación, Desarrollo y Sintonización de Aplicaciones Paralelas [2, 3]), que integra la automatización de los procesos de desarrollo y sintonización de aplicaciones paralelas basadas en diferentes *problem solvers*. En este proyecto particular, se propone trabajar sobre el problem solver Evolución Diferencial, y actualmente se espera obtener financiamiento a través del proyecto SIUTNME0004819 (UTN) el cual se encuentra en evaluación.

1. INTRODUCCIÓN

La Evolución Diferencial (ED) [7] constituye una metaheurística poblacional [4] muy utilizada para la resolución de problemas de optimización, basada en diferencia de vectores para el tratamiento de la población. Cada paso evolutivo incluye tres fases, a saber: la mutación, el cruzamiento y la selección/reemplazo. El operador de mutación es el que efectivamente aplica la diferencia de vectores entre los miembros de la población corriente para determinar el grado y la dirección de los individuos resultantes. Se han estudiado y analizado las diferentes posibilidades de paralelización de ED, tanto para potenciar la calidad de los resultados obtenidos como la velocidad de procesamiento del método, lo cual ha incluido la revisión del estado del arte en la materia. En consecuencia, hemos considerado tres posibilidades de paralelización, basadas

en diferentes variantes del modelo de islas con migración [8], e implementadas sobre un modelo Master/Worker con topología de anillo (a uno o dos niveles). Una de las posibilidades se denomina **modelo subpoblacional**, pues una única población es dividida por el master en subpoblaciones procesadas por cada isla o worker. Otra de las posibilidades, denominada **modelo de islas clásico**, en la que cada isla maneja su propia población completa. La tercera posibilidad se denomina **modelo jerárquico de islas**, y compone los dos modelos anteriores de forma jerárquica. El algoritmo maneja múltiples poblaciones, que son procesadas por islas independientes, dotadas cada una de varios workers que realizan los cálculos. Los resultados obtenidos [9] permitieron concluir que el modelo subpoblacional reduce el tiempo total de ejecución, en detrimento de la calidad de las soluciones, que disminuye dado el tamaño de las subpoblaciones; que usando el modelo de islas clásico los resultados mejoran notablemente en detrimento del tiempo de ejecución; y que en el modelo jerárquico de islas el algoritmo alcanza una relación de compromiso entre la calidad de las soluciones y el tiempo de ejecución. Los tres modelos han sido implementados, y asimismo, se ha trabajado en la calibración a diferentes niveles de todos ellos a fin de ajustar su comportamiento y mejorar la calidad general del funcionamiento. Para ello fue diseñado un conjunto de experimentos a fin de ejecutar diferentes combinaciones de parámetros y problemas para cada prototipo desarrollado. Los resultados obtenidos han sido analizados a fin de establecer la incidencia de cada parámetro, lo cual resulta de suma importancia para la determinación de los parámetros factibles de ser sintonizados estática y/o dinámicamente y así

identificar los problemas de rendimiento típicos de las aplicaciones paralelas, y específicos de la ED. Uno de los objetivos más importantes dentro de este proyecto consiste en encapsular todo este conocimiento, mediante el modelado del algoritmo ED y la definición de un modelo de rendimiento del mismo, que provea los elementos necesarios para abordar la sintonización del método de forma automática y dinámica. Que la sintonización sea automática, significa que las técnicas, capacidades, y conocimiento de rendimiento estarán incorporadas en la propia aplicación de forma automática, y como consecuencia el usuario no tendrá necesidad de responsabilizarse de la sintonización, sino que resultará una capacidad incluida en la aplicación de forma transparente para el usuario, lista para ser utilizada. Ello representa una gran ventaja dado que le da acceso al usuario a la reutilización de conocimiento experto en el análisis y resolución de problemas de rendimiento, sin necesidad de involucrarse en ninguna etapa del complejo proceso de sintonización.

Previo el avance en tal sentido se ha seleccionado un caso de aplicación particular de ED: la predicción de incendios forestales, trabajando sobre la representación de la información y el procesamiento de la misma, y diseñando una serie de experimentos que permitan analizar el impacto de los diferentes parámetros en la calidad de la ejecución del caso de aplicación. Por otro lado, se realizó un estudio comparativo [6] entre la predicción de incendios utilizando esta versión paralela de ED frente a otros métodos ya conocidos de predicción de incendios forestales [1, 5], para así evaluar la efectividad y eficiencia del método, cuestión en la que se continúa trabajando, dado que dicha comparación

requiere además afrontar la sintonización del método para mejorar su efectividad.

2. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

Esta línea de investigación tiene como objetivo general el desarrollo de un Entorno de Especificación, Desarrollo y Sintonización de Aplicaciones Paralelas (EEDSAP). El mismo ha sido concebido como un entorno integral para el desarrollo y la sintonización de aplicaciones paralelas, que hace transparente los aspectos relacionados con el diseño de aplicaciones paralelas y el proceso de sintonización automática y dinámica de las mismas, por medio de la utilización de esqueletos, plantillas e instrumentación de rendimiento. Estas características lo hacen especialmente útil para usuarios no expertos en paralelismo, sintonización, y hasta en programación, dado que el entorno ofrece una interfaz declarativa, por lo que no se requieren conocimientos específicos de programación por parte del usuario. En términos de intervención del usuario en el ciclo de vida de la aplicación, EEDSAP propone una primera fase cooperativa en la que el usuario describe el problema, una segunda fase automática en la que se genera el código de la aplicación y se lo dota de capacidades para su sintonización a partir de la interpretación de un esqueleto con la información provista por el usuario en la especificación; finalmente, en una tercera fase dinámica (es decir durante la ejecución) los parámetros de ejecución de la aplicación son sintonizados o ajustados a fin de alcanzar un rendimiento superior y una utilización más eficiente de los recursos computacionales involucrados, de acuerdo al estado del entorno de

ejecución. En este proyecto se dará tratamiento al *problem solver* Evolución Diferencial a fin de extender esta aproximación y así ampliar el uso y utilidad de EEDSAP.

3. RESULTADOS ESPERADOS

El principal aporte que se espera alcanzar es la dotación del método evolución diferencial con la capacidad de ajustar automática y dinámicamente aquellos parámetros de funcionamiento que permitan alcanzar una ejecución más eficiente. Tal capacidad será heredada por todas aquellas aplicaciones que se generen automáticamente a partir de él, dado que tanto el método ED como su modelo de rendimiento serán encapsulados en un módulo de software que permita generar aplicaciones paralelas sintonizables basadas en ED, haciendo transparentes tanto los conceptos de paralelismo como los de sintonización a la hora de desarrollar la aplicación. En otras palabras, el aporte general que se espera consiste en ofrecer una herramienta que requiera del usuario mínimos conocimientos relacionados con el paradigma paralelo, lenguajes de programación, lenguajes de especificación, mientras que ofrezca un entorno amigable e integral tanto para el desarrollo como para la sintonización automática de aplicaciones paralelas basadas en ED. Ello constituirá una contribución importante también para la línea de investigación que enmarca a este proyecto, cuyo objetivo se centra en el desarrollo de un entorno que integre el desarrollo automático con la sintonización, donde se espera que ambos procesos resulten transparentes al usuario, a la hora de crear y utilizar aplicaciones paralelas, y como parte del cual se incorporará ED como un nuevo *problem*

solver de desarrollo y sintonización. Tales características resultan primordiales en una herramienta pensada para usuarios no expertos, tanto en el campo del paralelismo, como en el de la sintonización, e incluso de los propios métodos de resolución (sean algoritmos genéticos, evolución diferencial, o cualquier otro que se integre a EEDSAP). Otro aporte importante que se espera alcanzar es la definición de un modelo de rendimiento de ED que encapsule la representación de las características del funcionamiento del *problem solver*, y permita analizar los problemas de rendimiento que manifiesta.

4. FORMACIÓN DE RECURSOS HUMANOS

La temática propuesta por este proyecto permite continuar con la formación de los distintos integrantes del grupo de trabajo, de forma complementaria a la formación adquirida hasta el momento, ahora en el campo del análisis y sintonización de software basado en evolución diferencial. En el caso particular de la Lic. Tardivo y el Ing. Méndez, su participación en el proyecto, el estudio, experimentación y puesta en marcha de los procesos de análisis y sintonización de metaheurísticas como la evolución diferencial, será fundamental como elemento para su formación en el área de sintonización de aplicaciones paralelas, aplicable a sus respectivos planes de trabajo. Ambos estudiantes cursan el Doctorado en Ciencias de la Computación de la Universidad Nacional de San Luis, y revisten como becarios doctorales de CONICET. Asimismo, el grupo de trabajo siempre está abierto a la incorporación de nuevos integrantes (de grado o postgrado) que deseen

familiarizarse con las temáticas que aquí se describen.

5. BIBLIOGRAFÍA

- [1] Bianchini, G., Caymes Scutari, P., Méndez Garabetti, M., Evolutionary-Statistical System: a Parallel Method for Improving Forest Fire Spread Prediction, *Journal of Computational Science (JOCS, Elsevier)*. Vol. 6, pp. 58-66, 2015.
DOI: 10.1016/j.jocs.2014.12.001
- [2] P. Caymes Scutari, G. Bianchini, Environment for the Automatic Development and Tuning of Parallel Genetic Algorithms. *Proceedings of 40 JAIIO at the High-Performance Computing Symposium (HPC 2011, ISSN: 1851-9326) (2011) 17 a 20.*
- [3] P. Caymes Scutari, G. Bianchini, A. Sikora, T. Margalef, Environment for Automatic Development and Tuning of Parallel Applications. *Proceedings of the 2016 International Conference on High Performance Computing & Simulation (HPCS 2016) ISBN: 978-1-5090-2088-1/16/\$31.00 ©2016 IEEE pp. 743-750. DOI: 10.1109/HPCSim.2016.7568409.*
- [4] Goldberg, D.E., Genetic and evolutionary algorithms, *Come of age*, *Communications of the ACM*, 37 (3) pp. 113119, 1994.
- [5] Méndez-Garabetti, M., Bianchini, G., Tardivo, M. L., & Caymes-Scutari, P., Comparative analysis of performance and quality of prediction between ESS and ESS-IM. *Electronic Notes in Theoretical Computer Science*, 314, 45–60, 2015. DOI: 10.1016/j.entcs.2015.05.004
- [6] Tardivo, M.L., Caymes Scutari, P., Bianchini, G., Méndez Garabetti, M., Cencerrado, A., Cortés, A., A comparative study of evolutionary statistical methods for uncertainty reduction in forest fire propagation prediction. *Procedia Computer Science*, 108, 2018–2027, 2017.
DOI: 10.1016/j.procs.2017.05.252
- [7] R. Storn, K. Price, *Differential Evolution - A simple and efficient adaptive scheme for global optimization over continuous spaces*, Technical Report TR-95-012, Berkeley, CA (1995).
- [8] E.G. Talbi, *Metaheuristics: From Design to Implementation*, John Wiley & Sons, Hoboken, New Jersey (2009).
- [9] M.L. Tardivo, P. Caymes-Scutari, M. Méndez-Garabetti, G. Bianchini, Parameters Calibration for Parallel Differential Evolution based on Islands. *XIII Workshop Procesamiento Distribuido y Paralelo, CACIC 2013*. 296-305, 2013.