

IQR: Una medida estadística como modelo para la sintonización computacional

Caymes-Scutari Paola^{1,2}, Tardivo María Laura^{1,3}, Bianchini Germán¹, Méndez-Garabetti Miguel¹

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

RESUMEN

El Rango Intercuartil (o métrica IQR) es una medida estadística que cuantifica la dispersión de la muestra considerada, es decir, la variabilidad de la distribución de los elementos muestrales en base a algún parámetro de interés. En este proyecto, centrado en métodos que como muestra del espacio de búsqueda consideran poblaciones de individuos (muestra de elementos candidatos) se propone la utilización del rango intercuartil de los valores de aptitud de los individuos como un indicador de la tendencia a estancarse que tiene el algoritmo. El caso de estudio se refiere a la predicción de incendios forestales. Por lo tanto, la función de aptitud cuantifica el grado de coincidencia entre la predicción arrojada por cierto individuo y el incendio real. El IQR actuaría como un indicador de estancamiento y/o convergencia prematura y sería de utilidad para tomar decisiones sobre cuándo la predicción alcanzada es suficientemente buena, o

bien ha alcanzado un cierto tope de calidad que no podrá mejorarse, y por consiguiente sea recomendable dar paso a una nueva población y una nueva generación. En resumen, se propone un modelo basado en el IQR para sintonizar de forma automática y dinámica el parámetro que regula la cantidad de generaciones del proceso evolutivo, a fin de evitar estancamiento y convergencia prematura.

Palabras clave: Rango Intercuartil, Evolución Diferencial, Algoritmos Evolutivos, Sintonización, Reducción de Incertidumbre.

CONTEXTO

El proyecto SIUTNME0004819 se encuentra en ejecución desde mayo de 2018, llevándose a cabo en el marco del LICPaD (Laboratorio de Investigación en Cómputo Paralelo/Distribuido) dentro del ámbito de la UTN-FRM. El mismo se encuentra enmarcado en el desarrollo de

un entorno denominado EDDSAP (Entorno de Especificación, Desarrollo y Sintonización de Aplicaciones Paralelas [1, 2]), 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 abordar la sintonización de los problemas de rendimiento que pueda presentar, mediante la consideración del rango intercuartil (IQR) como métrica de dispersión.

1. INTRODUCCIÓN

La predicción del comportamiento de un incendio forestal consiste en determinar cómo será la propagación del fuego sobre el terreno en un instante de tiempo futuro. Generalmente, los métodos de predicción implementan modelos que describen el comportamiento del fuego, y utilizan como dato de entrada un grupo de variables representando aquellos factores que condicionan la propagación. Entre ellas encontramos la velocidad y dirección del viento, la pendiente del terreno, el tipo de material combustible, la humedad de dicho material, etc. Lamentablemente, no es posible contar con los valores exactos para estos factores, debido a la imposibilidad de dotar todo el terreno forestal con instrumentos de medición y, sobre todo, debido a que algunos cambian dinámicamente durante el desarrollo del incendio [3]. Durante los últimos años se ha desarrollado una serie de métodos que pretenden reducir el impacto negativo que causa esta falta de conocimiento acerca del valor de las variables. Tal es el caso del método ESSIM-DE (Evolutionary Statistical Method with Island Model and Differential Evolution), el cual utiliza

Estadística para obtener el patrón de comportamiento de la línea de fuego, el algoritmo evolutivo Evolución Diferencial [4], para orientar la búsqueda hacia mejores soluciones, y Cómputo Paralelo a través de un modelo de islas y múltiples poblaciones, con dos jerarquías de procesos que cooperan en la obtención de las predicciones, lo que permite explorar un amplio espacio de búsqueda y a la vez acelerar el proceso de predicción [5]. ESSIM-DE ha sido estudiado en el último tiempo con el fin de identificar aquellos aspectos que puedan ser factores limitantes de mejor rendimiento, en términos de calidad de las predicciones obtenidas y tiempo de respuesta. En este sentido, se propone dotar a ESSIM-DE con capacidades para su sintonización dinámica, lo que habilitará un mejor aprovechamiento de los recursos al adaptar durante la propia ejecución ciertos parámetros dependientes del problema o del estado del sistema. Dicho proceso de sintonización se logra a partir de cuatro fases que sucesivamente actúan sobre la aplicación [6]: *instrumentación* para anotar la aplicación con las métricas necesarias, *monitorización* para recolectar y clasificar las métricas, *análisis* para evaluar el estado del sistema en función de las métricas, y *sintonización* en sí misma para concretar los ajustes necesarios en la aplicación a fin de adaptar su comportamiento.

2. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

Esta temática se enmarca en la línea de investigación enfocada en el proceso de sintonización automática y dinámica de aplicaciones paralelas. El objetivo es detectar y modelar los problemas de rendimiento que pudiese traer

intrínsecamente aparejados el algoritmo de base considerado en la aplicación, a fin de darles tratamiento para paliar sus efectos negativos y obtener una ejecución más eficiente desde el punto de vista funcional, energético y de precisión en los resultados obtenidos. En este proyecto, se considera un método de predicción basado en evolución diferencial, y para abordar el proceso de sintonización hemos considerado dos posibles problemas asociados a su proceso evolutivo: la convergencia prematura y el estancamiento. La convergencia prematura es la situación en la cual la población converge a un óptimo local, debido a la pérdida de diversidad. Por su parte, el estancamiento es la situación en la cual el optimizador no es capaz de generar una nueva solución mejor que la anterior, aun cuando la población no hubiese convergido. Esto significa que, incluso cuando la población presente cierta diversidad, el optimizador es incapaz de encontrar mejores soluciones [7].

Así es que, para estos problemas, la sintonización se propone desde diversas aristas. Una de ellas consiste en la incorporación de un nuevo operador de reinicio (denominado r) de la población al inicio de cada paso de predicción, a fin de evitar el estancamiento global del proceso de búsqueda. Otra de las estrategias de sintonización propone la utilización de algún criterio para detectar, en cada generación, cuándo la búsqueda ha convergido lo suficiente, y por lo tanto poder interrumpir la ejecución del algoritmo a fin de obtener un resultado útil en un tiempo de ejecución menor. Para ello se propone analizar la tendencia decreciente de la distribución de la población a lo largo del proceso evolutivo, a fin de ajustar el rendimiento del método de acuerdo con el contexto de ejecución corriente, así como también al

tipo de incendio considerado y a los diferentes parámetros que guían el proceso evolutivo en sí. Si bien en este proyecto el modelo de sintonización se aplica a un sistema de predicción de incendios, su utilización puede generalizarse para mejorar la performance de otras metaheurísticas, métodos de predicción y/o casos de estudio. El análisis a realizarse sobre las características de la población corriente ha de tener en cuenta fundamentalmente la diversidad poblacional, que puede modelarse y valorarse por medio de la métrica IQR. Por ello, se propone computar el Rango Intercuartil de los valores de aptitud de los individuos [8] a fin de cuantificar la dispersión de la población, es decir, la variabilidad de la distribución de la población en base a la aptitud de cada individuo, y así determinar la tendencia a la convergencia y/o el estancamiento.

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. Dicho objetivo será alcanzado a partir de otro aporte más particular, centrado en la definición de un modelo de rendimiento de ED basado en el rango intercuartil (IQR) que presenta la población. A su vez, 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, su participación en el proyecto, el estudio, experimentación y puesta en marcha de los procesos de análisis y sintonización de la evolución diferencial, será fundamental como elemento para su formación en el área de sintonización de aplicaciones paralelas, aplicable a su plan de trabajo. En el caso del Mg. Ing. Méndez Garabetti la temática también resulta transversal a su plan de tesis doctoral. Ambos estudiantes cursan el Doctorado en Ciencias de la Computación de la Universidad Nacional de San Luis. 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] 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.
- [2] 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.

- [3] Bianchini, G., Denham, M., Cortés, A., Margalef, T., & Luque, E., Wildland fire growth prediction method based on Multiple Overlapping Solution. *Journal of Computational Science*, 1(4), 229–237, 2010.
DOI: 10.1016/j.jocs.2010.07.005
- [4] Price K., Storn R., Lampinen J. (2005) *Differential Evolution - A practical approach to global optimization*. Springer-Verlag New York, Inc.
- [5] Talbi, E. (2009) *Metaheuristics: From Design to Implementation*. John Wiley & Sons, Hoboken, New Jersey
- [6] Naono, K., Teranishi, K., Cavazos, J. y Suda, R. (2010) *Software Automatic Tuning: From Concepts to State-of-the-Art Results*, Springer, New York.
- [7] Lampinen, J. y Zelinka, I. (2000), “On the Stagnation of the Differential Evolution algorithm”, *I.C. Soft Computing*, pp. 76-83.
- [8] Healey, J. F. (2007) *The Essentials of Statistics: A Tool for Social Research*, Thomson/Wadsworth.