

Diseño Óptimo de Parques Eólicos con Metaheurísticas Poblacionales y Técnicas de Minería de Datos utilizando Procesamiento Paralelo

de San Pedro M.* , Pandolfi D.* , Lasso M.* , Villagra A.* , Lorenzetti D.# ,
Valdez J.* , Varas V.* , Vidal P.* , Bilbao M.*

(*) Laboratorio de Tecnologías Emergentes (LabTEm)

(#) Laboratorio de Energías Renovables

Unidad Académica Caleta Olivia

Universidad Nacional de La Patagonia Austral

e-mail: {edesanpedro, dpandolfi, mlasso, avillagra, dloren, jcvaldez, vvaras}@uaco.unpa.edu.ar;
{martin.bilbao, pablo.vidal.20}@gmail.com

Resumen

El trabajo que se viene realizando en este grupo de investigación, está enfocado al estudio de tres líneas bien definidas en donde se vienen realizando diferentes trabajos.

En cuanto a *Metaheurísticas*, el grupo ha adquirido en los últimos años una importante experiencia, siempre con el fin de producir versiones mejoradas de las mismas respecto a sus capacidades explorativas para su aplicación en muchos campos del mundo real.

Asimismo, el campo de *Minería de Datos* fue incorporado al grupo como objeto de estudio y sus posibles aplicaciones en el último proyecto presentado y actualmente en ejecución cuyo principal objetivo es el uso de metaheurísticas en el campo de minería de datos para mejorar el desempeño de las técnicas de minería de datos o como técnicas de minería de datos en sí.

Finalmente, en cuanto a *Energía Eólica* y la problemática de la instalación de parques eólicos, el grupo ha generado una fluida interacción con integrantes de un grupo de investigación especializados en energías renovables, considerando que los objetivos propuestos pueden ser alcanzados.

Palabras clave: *Minería de datos, Energía Eólica, Modelos descriptivos y predictivos, Metaheurísticas, Problemas de Asignación*

Contexto

Este trabajo es desarrollado por el grupo de investigación que actualmente tiene una trayectoria y amplia experiencia demostrada en metaheurísticas, a través de la sostenida producción científica desde sus inicios en el año 1998, en publicaciones nacionales ([Villagra et al, 2005a] [Villagra et al, 2005b] [Villagra et al, 2004b] [Pandolfi et al, 2003a] [de San Pedro et al, 2003a] [Lasso et al, 2003a] [Lasso et al, 2003b] [de San Pedro et al, 2002a] [Lasso et al, 2002a] [Pandolfi et al, 2002c]) e internacionales ([de San Pedro et al, 2005] [Villagra et al, 2004a] [Lasso et al, 2004] [de San Pedro et al, 2004] [Lasso et al, 2003c] [de San Pedro et al, 2003c] [Pandolfi et al, 2003b] [Pandolfi et al, 2002a] [Pandolfi et al, 2002d]) en eventos de reconocida jerarquía, como también varios artículos en revistas [Pandolfi et al, 2004] [Pandolfi et al, 2002b].

Estas líneas de Investigación forman parte del proyecto “*Minería de Datos y Técnicas Metaheurísticas: su aplicabilidad al problema de explotación eficiente de energía eólica a través de la obtención de modelos aproximados y distribución óptima de máquinas eólicas*” del Laboratorio de Tecnologías Emergente (LabTEm) en el marco del programa de Investigación en Ciencia y Tecnología de la Universidad Nacional de la Patagonia Austral.

Introducción

La energía eólica es una de las energías alternativas más importantes del mundo. Se trata de una energía económica, libre y limpia, y que en la actualidad puede competir con otros tipos de energía. El objetivo principal es producir el máximo de energía al mismo tiempo que se reduce el costo total del parque eólico.

Más precisamente, un parque eólico es una agrupación de máquinas eólicas o aerogeneradores, usadas para producir energía eléctrica y las cuales se distribuyen sobre una región generalmente extensa. La posición de cada uno de estos aerogeneradores es una decisión estratégica de reducir el efecto estela [Manwell, 2003] a los fines de maximizar la energía producida.

Con este objetivo se ha trabajado en la obtención de una configuración óptima del parque eólico usando las condiciones del viento y el terreno dado por el medio ambiente.

Existe un trabajo realizado con datos reales obtenidos de una distribución del viento en la ciudad de Comodoro Rivadavia tomados en 2008 [<http://www.eolica.com.ar/>]. Para esto se utilizaron algoritmos eficaces, que debieron ser evaluados por primera vez antes de su utilización [Bilbao, 2010a] [Bilbao, 2010b].

El Simulated Annealing [Bilbao, 2009a] y los Algoritmos Genéticos Distribuidos [Huang, 2007] son dos metaheurísticas que han sido utilizados para resolver este tipo de problema. En un trabajo anterior hemos utilizado CHC y GPSO considerando constante el viento del Norte [Bilbao, 2009b]. Ahora, comparamos diferentes escenarios con viento constante del Oeste y un cuarto escenario usando la distribución real del viento. Analizamos la mejor configuración del parque encontrado, el valor de *fitness*, la energía producida, la eficiencia, el rendimiento de los algoritmos en términos de su ejecución, tiempo y número de evaluaciones necesarios para obtener la mejor solución

La minería de datos constituye el núcleo del análisis inteligente de los datos y ha recibido un gran impulso en los últimos tiempos motivado por distintas causas: i) el desarrollo de algoritmos eficientes y robustos para el procesamiento de grandes volúmenes de datos, ii) un poder computacional más barato que permite utilizar métodos computacionalmente

intensivos, y iii) las ventajas comerciales y científicas que han brindado este tipo de técnicas en las más diversas áreas.

Es importante diferenciar en la minería de datos, el tipo de tareas que se suelen abordar y las técnicas utilizadas en cada caso. Como ejemplos de tareas generales, se pueden mencionar el aprendizaje de conceptos, clasificación, categorización, regresión, agrupamiento (o clustering), correlaciones y análisis de asociación. Estas tareas pueden ser abordadas mediante distintos métodos o técnicas que suelen adaptarse mejor de acuerdo a la tarea sobre la cual se trabajará. Entre las técnicas más conocidas se puede mencionar el aprendizaje de reglas de clasificación, reglas de asociación, reglas relacionales, reglas difusas, árboles de decisión (y regresión), ecuaciones de regresión, redes neuronales, metaheurísticas, etc. [Witten 1999], [Witten 2005], [Orallo et al. 2003].

En el contexto de construcción y mantenimiento de plantas productoras de energía, es indudable que a medida que los costos de los equipos para capturar de diversos datos disminuyen y por ende se expande su utilización, la cantidad de datos recolectados se incrementa proporcionalmente. En la actualidad existe un término (Energy Data Mining), que identifica a todos aquellos procesos de búsqueda automática en grandes volúmenes de datos, derivados de observaciones vinculadas a plantas productoras de energía, que intentan descubrir patrones ocultos, correlaciones entre diferentes atributos y modos operacionales sostenidos. Por lo tanto, tareas típicas de minería de datos, tal como clustering, regresión, clasificación, reglas de asociación, etc., pueden ser aplicadas sobre esos datos que ayuden a la toma de decisiones para un adecuado diseño y/o funcionamiento de una planta productora de energía. En el caso de la energía eólica y desde la perspectiva del proyecto propuesto, estamos ante al siguiente situación: por un lado, es importante tratar de ajustar los modelos del comportamiento del viento en ciertas regiones de la Patagonia, a través del análisis de datos que permitan, describir o explicar el comportamiento de los datos; y por otro lado, poder realizar ciertas predicciones para mejorar el aprovechamiento de los recursos disponibles cuando la planta se encuentre en proceso de producción.

(<http://www.osdpd.noaa.gov/ml/index.html>).

Sin embargo, existen otras decisiones importantes antes de poner en funcionamiento un parque eólico de manera tal que pueda aprovechar al máximo las condiciones climáticas (comportamiento del viento en la zona elegida). En este caso, es de vital importancia la distribución de las máquinas eólicas, ya que este tipo de decisiones no son tan simples. Hay un aspecto de fundamental importancia, que nuestra propuesta no considera, y que está vinculado al tipo y diseño de los aerogeneradores a ser usados en la instalación.

En otro aspecto, y considerando el análisis de diferentes metaheurísticas y a los fines de lograr mejores resultados al momento de analizar los resultados, es la necesidad de realizar procesos en paralelo. El interés en la computación paralela es un tema recurrente que ha sido estudiado una y otra vez desde la aparición de los primeros ordenadores. En los últimos años, ha crecido el interés de los investigadores en el uso de las unidades de proceso gráfico (en adelante GPUs, de Graphics Processing Units) de bajo costo para las aplicaciones que requieren intensiva computación paralela. Esto se debe en gran medida a la capacidad de estos dispositivos para resolver problemas en forma paralelizable mucho más rápido que los tradicionales procesadores secuenciales.

Nuestra investigación se centra en utilizar la enorme potencia de cálculo de las GPUs para desarrollar aplicaciones fuera del entorno gráfico, esto es conocido como programación de propósito general sobre GPUs (GPGPU). En este contexto, nuestras actividades se enfocan en estudiar el comportamiento de diversos algoritmos de optimización. Para ello, se realiza una formulación y posterior codificación de los algoritmos analizados en el modelo de programación paralela llamado CUDA, el cual es empleado para desarrollar sobre las GPUs. CUDA se basa en un modelo de programación paralela que funciona como una extensión del lenguaje C.

En síntesis, el presente proyecto tiene dos objetivos principales, uno de ellos, destinado a la aplicación de técnicas de minería de datos para el análisis inteligente de datos obtenidos a través de diversas mediciones in situ (provistas por diferentes fuentes), que permitan inferir modelos descriptivos y predictivos del

comportamiento del viento. La otra parte, orientada a la investigación y desarrollo de metaheurísticas (versiones seriales y paralelas), para resolver uno de los problemas claves (distribución) asociados al proceso de instalación de aerogeneradores en un parque eólico.

Lineas de Investigación y desarrollo

Teniendo en cuenta los aspectos considerados anteriormente, se pueden desprender las siguientes líneas temáticas específicas:

A) Aplicación de técnicas metaheurísticas para el diseño óptimo de parques eólicos, que permita maximizar la energía producida.

B) Uso del procesamiento paralelo de las Metaheurísticas desarrolladas, a los fines de acelerar el proceso de búsqueda y mejorar la calidad de las soluciones.

C) Utilización de técnicas de Minería de Datos, aplicadas a tareas descriptivas y predictivas.

Resultados y Objetivos

Se han realizado los siguientes trabajos, considerando las líneas de investigación mencionadas.

Con respecto a la aplicación de diferentes técnicas metaheurísticas para el diseño óptimo de parques eólicos, se están realizando las siguientes tareas:

- Aplicación de las técnicas sobre terrenos irregulares, mejorando las funciones de penalidad para las soluciones y las influencias de diferentes valores de k de weibull en el parque eólico.
- Se está optimizando la distribución de weibull logrando una mejor aproximación a los datos reales obtenidos en el 2010.
- Incorporación de nuevos aerogeneradores al diseño de parques eólicos, con características diferentes para que el sistema elija

convenientemente en función a los objetivos propuestos.

- Diseño de nuevos algoritmos con diferentes representaciones para estudiar el comportamiento y las mejoras que se pueden obtener.

En relación al estudio de la computación paralela, nuestros primeros pasos se han dirigido al estudio de los algoritmos genéticos celulares, los cuales se presentan como una excelente opción para evaluar el desempeño de las GPUs frente a las CPUs, midiendo la eficiencia como también la eficacia de los algoritmos implementados en ambas arquitecturas [Vidal, 2010a] [Vidal, 2010b].

Actualmente nos encontramos dedicados al diseño y fusión entre las ideas de computación sistólica, algoritmos de optimización y metaheurísticas paralelas. Nuestra primera contribución ha sido considerar a la GPU como una herramienta para desplegar nuestro algoritmo llamado SNS (Systolic Neighborhood Search) buscando resolver un problema de optimización de una manera sistólica [Vidal, 2011].

Como trabajos futuros y como continuación de todo este trabajo se planea desarrollar una librería de algoritmos de optimización para GPUs y asimismo continuar el estudio del SNS sobre otros dominios de problemas.

Por otra parte, se prevé el estudio de diferentes objetivos a optimizar, mejoras en las funciones de potencia, costos, optimización del VAN (Valor Anual Neto).

Además, nuevos estudios de algoritmos paralelos sobre memoria compartida y memoria distribuida, estudio de nuevos mecanismos de migración y selección de individuos.

Por ultimo, enfoques del estudio de optimización de parques eólicos multiobjetivo (optimizando más de un objetivo a la vez).

Formacion de Recursos Humanos

Dos de los integrantes, están actualmente realizando sus estudios de postgrado (Doctorado), que iniciaron en Septiembre de 2010, en la Universidad de Málaga bajo la dirección del Dr. Enrique Alba, para realizar

estudios sobre Modelos y Algoritmos Paralelos en el ámbito de Metaheurísticas.

Dos integrantes, han comenzado a definir su plan de tesis en temas afines a los propuestos en el proyecto, y han comenzado a realizar cursos de postgrado como parte inicial de sus tareas para definir su plan.

Dos becarios de investigación se han incorporado al grupo a fines de 2010.

Referencias

- [Bilbao 2009a] M.Bilbao, E. Alba Simulated Annealing for Optimization of Wind Farm Annual Profit, LINDI 2009, Linz, Austria.
- [Bilbao 2009b] M.Bilbao, E. Alba GA and PSO Applied to Wind Energy Optimization, CACIC 2009, Jujuy, Argentina.
- [Bilbao 2010a] M.Bilbao, E. Alba CHC and SA Applied to Wind Energy Optimization Using Real Data, CEC 2010, Barcelona, España.
- [Bilbao 2010b] M.Bilbao, E. Alba Designing Optimal Wind Farm With Parallel CHC, ALIO-INFORMS Join International Meeting 2010, Buenos Aires, Argentina.
- [de San Pedro et al, 2002a] de San Pedro M., Pandolfi D., Villagra A., Lasso M., Vilanova G, Gallard R. - Adding problem-specific knowledge in evolutionary algorithms to solve W-T scheduling problems - CACIC 2002, Buenos Aires, 2002.
- [de San Pedro et al, 2003a] de San Pedro M., Lasso M., Villagra A., Pandolfi D., Gallard R. - Influence of Crossover Operators in Evolutionary Scheduling Under Multirecombined Schemes - CACIC 2003, La Plata, 2003.
- [de San Pedro et al, 2003c] de San Pedro M., Villagra A., Lasso M., Pandolfi D., Díaz Vivar M., Gallard R. - Solutions for the Weighted Number of Tardy Jobs in Single Machine Environments via Evolutionary Algorithms - CSITeA03, Rio de Janeiro, 2003.
- [de San Pedro et al, 2004] de San Pedro M., Pandolfi D., Lasso M., Villagra A., Gallard R. - Effect of Crossover Operators under Multirecombination: Weighted Tardiness, a Test Case – CEC '04 – Portland, U.S.A., 2004.

- [de San Pedro et al, 2005] de San Pedro M., Pandolfi D., Lasso M., Villagra A. - Dynamic Scheduling Approaches to solve Single Machine Problems, ASC 2005, Benidorm, España, 2005.
- [Huang, 2007] H.S. Huang, “Distributed Genetic Algorithm for Optimization of Wind Farm Annual Profits”. Intelligent Systems Applications to Power Systems, 2007. ISAP 2007. International Conference on Volume, Issue, 5-8 Nov. 2007 Page(s):1 - 6.
- [Lasso et al, 2002a] Lasso M., Pandolfi D., de San Pedro M., Villagra A., Vilanova G, Gallard R. - Algorithms to solve the dynamic weighted tardiness problem - CACIC 2002, Buenos Aires, 2002.
- [Lasso et al, 2003a] Lasso M., de San Pedro M., Villagra A., Pandolfi D., Gallard R. - Solutions to the Dynamic Average Tardiness Problem in Single Machine Environments - CACIC 2003, La Plata, 2003.
- [Lasso et al, 2003b] Lasso M., Pandolfi D., de San Pedro M., Villagra A., Vilanova G, Gallard R. - Heuristics for partial and total dynamic w-t problems In single machine environments - WICC'03, Tandil, 2003.
- [Lasso et al, 2003c] Lasso M., Pandolfi D., de San Pedro M., Villagra A., Gallard R. - Heuristics to Solve Dynamic W-T problems in Single Machine Environments - CSITeA03, Rio de Janeiro, 2003.
- [Lasso et al, 2004] Lasso M., Pandolfi D., de San Pedro M., Villagra A., Gallard R. - Solving Dynamic Tardiness Problems in Single Machine Environments” – CEC '04 – Portland, U.S.A., 2004.
- [Manwell, 2003] J.F. Manwell, J.G. McGowan, and A.L. Rogers, Wind Energy Explained-Theory, Design and Application, 1st ed. , Reprint with correction, John Wiley & Sons Ltd. ,2003, p.384 & p.44.
- [Orallo et al. 2003] C. Ferri Ramírez; J. Hernández Orallo; M.J. Ramírez Quintana Introducción A La Minería De Datos (Pearson Educación), 2003.
- [Pandolfi et al, 2002a] Pandolfi D., de San Pedro M., Villagra A., Vilanova G, Gallard R. - Multirecombining Random and Seeds with Studs in evolutionary algorithm to solve W-T Scheduling problems - CSITeA-02, Foz Iguazú, Brasil, 2002.
- [Pandolfi et al, 2002b] Pandolfi D., de San Pedro M., Villagra A., Vilanova G, Gallard R. - Stud Mating Immigrants in Evolutionary Algorithms to solve the Earliness-Tardiness Scheduling Problem - Cybernetics and Systems: An International Journal, Ed Taylor & Francis, pp 391-400, 2002 Vol 33 Number 4 June 2002 ISSN 0196-9722
- [Pandolfi et al, 2002c] Pandolfi D., de San Pedro M., Villagra A., Vilanova G, Gallard R. - Inserting Problem Specific-Knowledge in Multirecombined Evolutionary Algorithms - WICC'02, 2002.
- [Pandolfi et al, 2002d] Pandolfi D., de San Pedro M., Villagra A., Vilanova G, Gallard R. - A survey an multirecombined evolutionary approach for single machine scheduling – AEB'02, Merida, España, 2002.
- [Pandolfi et al, 2003a] Pandolfi D., Lasso M., de San Pedro M., Villagra A., Gallard R. - Knowledge Insertion: an Efficient Approach to Reduce Search Effort in Evolutionary Scheduling - CACIC 2003, La Plata, 2003.
- [Pandolfi et al, 2003b] Pandolfi D., Lasso M., de San Pedro M., Villagra A., Gallard R. - Evolutionary algorithms to solve average tardiness problems in single machine environments - CSITeA03, Rio de Janeiro, 2003.
- [Pandolfi et al, 2004] Pandolfi D., Lasso M., de San Pedro M., Villagra A., Gallard R. - Knowledge Insertion: an Efficient Approach to Reduce Search Effort in Evolutionary Scheduling - Journal of Computer Science & Technology, pp 109-114, Vol 4 Number 2 Agosto 2004 ISSN 1666-6038
- [Vidal 2010a] Vidal P. and Alba E., Cellular Genetic Algorithm on Graphic Processing Units, NCSO, May 2010, Granada, Spain
- [Vidal 2010b] Vidal P. and Alba E., A Multi-GPU Implementation of a Cellular Genetic Algorithm, CEC, July 2010, Barcelona, Spain
- [Vidal, 2011] Vidal P., Alba E. Systolic Optimization on GPU Platforms, 6 al 11 de Febrero 2011. Las Palmas de Gran Canaria, España, pag 294-295
- [Villagra et al, 2004a] Villagra A., de San Pedro M., Lasso M., Pandolfi D. - Multirecombined Evolutionary Algorithm inspired in the Selfish Gene Theory to face

- the Weighted Tardiness Scheduling Problem - Iberamia 2004 - Puebla, Mexico, 2004.
- [Villagra et al, 2004b] Villagra A., de San Pedro M., Lasso M., Pandolfi D. - Algoritmos Evolutivos inspirados en la Teoría del Gen Egoísta - WICC'04 – Neuquén, 2004.
- [Villagra et al, 2005a] Villagra A., de San Pedro M., Lasso M., Pandolfi D. – Algoritmo Evolutivo basado en el mecanismo de haplodiploidia para resolver el problema de planificación de weighted tardiness. CACIC 2005.
- [Villagra et al, 2005b] Villagra A., de San Pedro M., Lasso M., Pandolfi D. - Optimización guiada por Algoritmos Evolutivos Multirecombinados inspirados en la Teoría del Gen Egoísta para resolver Problemas de Weighted Tardiness – XI Reunión de Trabajo en Procesamiento de la Información y Control, 2005.
- [Witten 1999] I.H. Witten - Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, The Morgan Kaufmann Series in Data Management Systems, 1999.
- [Witten 2005] I.H. Witten - Data Mining: Practical Machine Learning Tools and Techniques, Second Edition, Morgan Kaufmann Series in Data Management Systems, 2005.