TEMA 1 - PLAN DE TRABAJO

TESIS DE LICENCIATURA EN CS DE LA COMPUTACIÓN

Diseño de algoritmos CAM^{b}D basados en metaheurísticas paralelas

Director: Dra. Nélida Beatriz Brignole

Ayudante: Ing. Mariana González Prieto

Introducción

El diseño de blends de gasolina y biocombustibles mediante CAM^{b}D (Computer Aided Mixture/blend Design) es una rama muy interesante pues proporciona un medio para estudiar el efecto de distintos compuestos oxigenados, productos de la transformación de la biomasa, en combustibles y de esta forma encontrar productos innovadores. CAM^{b}D (Karunanithi et al., 2005) es un concepto que implica identificar, dado un conjunto de compuestos determinado, una mezcla que presente las propiedades deseadas. Su objetivo general es diseñar el número óptimo, la identidad y la composición de los compuestos que forman una mezcla que cumpla con un criterio de diseño.

CAM^{b}D requiere el desarrollo de tres herramientas fundamentales: una base de datos de compuestos químicos candidatos a agentes de mezclado, una librería de modelos para predecir propiedades de combustibles y por último, un optimizador de mezclas. Las primeras dos herramientas mencionadas ya se encuentran desarrolladas, con los modelos de predicción de propiedades ya escritos en subrutinas de Fortran®. La tercera constituye una línea de investigación abierta. Un obstáculo en el desarrollo de herramientas para CAM^{b}D es la complejidad de los programas matemáticos necesarios para formular y resolver el problema. El diseño de una mezcla es un problema mixto entero no lineal (MINLP) en el cual aparecen dificultades relacionadas con la no linealidad de los modelos que describen las propiedades de las mezclas en conjunto con un amplio espacio de búsqueda. Alcanzar un óptimo es un desafío importante debido a la naturaleza combinatoria que aportan las variables binarias y a la alta no linealidad de las expresiones que relacionan composiciones y propiedades físicas.

El uso de metaheurísticas permite reducir significativamente la complejidad matemática del proceso de búsqueda; sin embargo, el tiempo de ejecución puede ser elevado en algunos problemas de interés real. Con la proliferación de plataformas

paralelas de cómputo eficientes, la implementación paralela de estas metaheurísticas surge de forma natural como una alternativa atractiva para acelerar la obtención de soluciones precisas a estos problemas.

Objetivo

El objetivo principal de este plan de trabajo consiste en el desarrollo de nuevos algoritmos basados en metaheurísticas. Se propone incorporar programación paralela en el algoritmo de diseño de blends.

Esta aplicación es un problema complejo del mundo real, donde las técnicas tradicionales han obtenido resultados satisfactorios cuando se utilizan modelos simplificados para el cálculo de propiedades de combustibles. Las metaheurísticas no han sido testeadas en este tipo de problemas. Este enfoque permitiría incorporar modelos más detallados y, mediante el uso de técnicas paralelas se podría resolver el costo de tiempo computacional que esto implica.

Para evitar trabajar con expresiones matemáticas complejas, se decide utilizar como método de búsqueda algoritmos genéticos. Estos a su vez permiten obtener un abanico de resultados, es decir varias mezclas solución que deben ser evaluados con más detalle. Además, los algoritmos genéticos permiten incorporar de forma directa conocimiento de la naturaleza química y comportamiento termodinámico de alto nivel, junto con estrategias razonables que guíen la búsqueda de forma eficiente.

El uso de técnicas metaheurísticas paralelas permitirá que el algoritmo se caracterice por su simplicidad, robustez y flexibilidad, con sustanciales mejoras de eficiencia con respecto al algoritmo secuencial existente. Existen numerosas alternativas para dicha implementación (Cantú-Paz, 1998).

Antecedentes

El etanol puede producirse a partir de azucares y celulosa. Este alcohol es el compuesto que se utiliza actualmente para mezclar con gasolinas sin embargo presenta algunas desventajas como combustible: posee un 70% del poder calorífico de la gasolina convencional, es miscible con agua y su mezcla con gasolinas hace que se incremente la presión de vapor del combustible resultante. Por este motivo, es interesante el estudio de otros compuestos oxigenados que se generan durante el procesamiento de biomasa. Por otro lado, la EPA (Enviromental Protection Agency) requiere que cualquier gasolina o mezcla gasolina-oxigenado vendida en los Estados Unidos cumpla con las especificaciones de la ASTM D4814-88. La norma fue desarrollada para asegurar que los requerimientos sean principalmente de

funcionamiento, no de concentración o formulación. Es decir, la norma no excluye ningún compuesto oxigenado, mientras sea la gasolina el componente principal del blend y se cumplan con los requerimientos de desempeño del combustible en el motor.

Partiendo de estas dos premisas, surge el interés en el desarrollo de un programa que permita seleccionar compuestos y proporciones de cada uno de forma tal de obtener un biocombustible que pueda reemplazar la gasolina convencional.

Pocos métodos han sido propuestos para resolver el problema de diseño de mezclas de más de dos componentes (Jonuzaj et al., 2015) , entre ellos se encuentran los trabajos de Klein et al. (1992) y el de Yunus et al. (2014). Los primeros proponen utilizar un algoritmo de optimización de dos lazos llamado Successive Regression and Linear Programming (SRLP). Yunus et al. (2014) proponen un enfoque sistemático de 4 pasos, llevando a cabo la optimización mediante un método de descomposición. Más recientemente en el trabajo de Jonuzaj et al. (2015) se propone utilizar la técnica de Generalized Disjunctive Programming (GDP). En este caso, una vez formulado el problema se traduce a un MINLP mediante técnicas de big M o convex hull.

En relación a la aplicación de metaheurísticas en el diseño de blends, se dispone de un algoritmo secuencial (González Prieto et al., 2015a y b) que utiliza el resolvedor de Algoritmos Genéticos disponible en Matlab®. El diseño consiste en seleccionar una cierta cantidad de compuestos de una lista de compuestos posibles y la cantidad de cada uno de ellos en la mezcla. Los posibles compuestos presentes en la mezcla se encuentran listados en función de su punto normal de ebullición (NBP) y se clasifican según el mismo en: compuestos muy livianos (ML), livianos (L), pesados (P) y muy pesados (MP). En la Figura 1se muestra una representación esquemática del individuo para una mezcla de 6 compuestos. Las variables y_i corresponden a la posición de cada compuesto en la lista de compuestos por lo que esta es una variable entera. Las variables x_i en cambio, son reales y corresponden a la cantidad de cada compuesto en la mezcla. La función de fitness elegida contabiliza la desviación entre las propiedades de cada posible mezcla solución (individuo) y las propiedades deseadas o fijadas por el usuario.

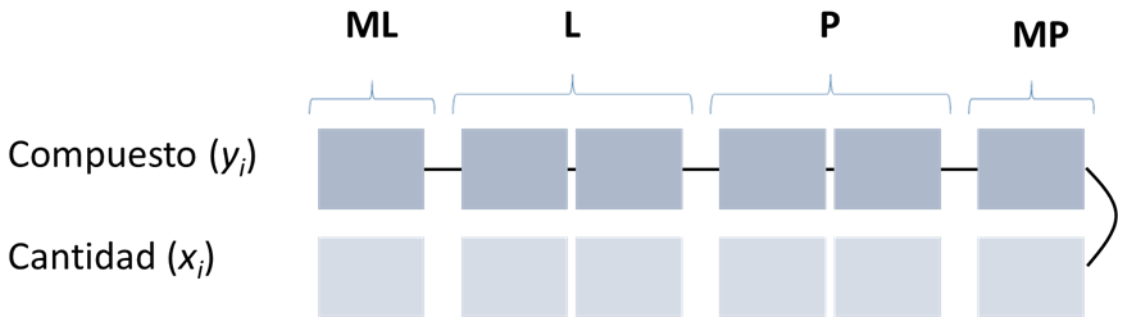


Figura 1: representación de un individuo

Actividades y metodología

El actual interés en la reducción del uso de combustibles fósiles y en la utilización de materias primas renovables genera numerosos desafíos para enfrentar y resolver en el campo de la optimización combinatorial relacionados con el diseño de blends. A continuación listamos las actividades propuestas a fin de lograr los objetivos enunciados anteriormente.

Act. 1: Diseño e implementación de un algoritmo con evaluación en paralelo de la función de fitness.

Act. 2: Implementación de un algoritmo basado en la gestión de poblaciones en paralelo.

Act 3: Verificación y comparación de desempeño de los algoritmos implementados en las Act. 1 y 2.

Act. 4: Exploración de nuevas alternativas de diseño.

Para la **Act. 1**, es posible habilitar los procesadores en Matlab® y utilizar la opción de evaluación de la función objetivo en paralelo disponible en el solver del algoritmo genético de Matlab®. Esta alternativa es sencilla de implementar y es prometedora con respecto a la reducción de tiempo de cómputo que se podría logar. Obedece a un esquema tipo maestro-esclavo, en el cual el maestro almacena la población y los esclavos hacen la evaluación del fitness. Esta es una operación muy común ya que la evaluación del fitness de cada individuo es una tarea independiente del resto de la población y no hay necesidad de comunicación durante esta etapa. Únicamente existe comunicación cuando los esclavos reciben el subset de individuos para evaluar y cuando los mismos retornan al maestro el valor de fitness de cada individuo. El algoritmo es de tipo sincrónico si se detiene y espera a recibir el valor del fitness de todos los individuos. Un esquema maestro-esclavo sincrónico tiene las

mismas propiedades que el algoritmo genético original diferenciándose únicamente en la velocidad de cómputo. Podría implementarse una ejecución asincrónica en la que el maestro no necesite detenerse a esperar todos los resultados de los esclavos, pero en este caso el desempeño del algoritmo genético será distinto del algoritmo original sin paralelizar.

En la **Act. 2** se propone trabajar con poblaciones en paralelo. Es importante tener en cuenta que esta implementación va a diferir del comportamiento del algoritmo original ya que cada procesador considerará sólo un subconjunto de individuos para la selección. Lo mismo ocurrirá con el cruzamiento ya que el mismo tendrá lugar dentro de la subpoblación. Lo interesante de esta alternativa es que las poblaciones son independientes lo que hace posible explorar distintas regiones del espacio de búsqueda. En Cantú-Paz (1998) se discute la idea de que en pequeñas poblaciones aisladas los rasgos favorables se propagan más rápido que en grandes poblaciones; sin embargo, cuando las poblaciones están aisladas, el crecimiento rápido del fitness se detiene en un valor más bajo que en una población de mayor tamaño. Lo equivalente es válido cuando se quiere minimizar una función objetivo. Para atacar este problema, se pueden intercambiar los mejores individuos de cada población en intervalos de tiempo o generaciones fijadas previamente. Esta taza de intercambio es una variable muy importante de diseño ya que permitirá evaluar cuál es el costo de comunicación que es aceptable afrontar para lograr un mejor desempeño del algoritmo.

La **Act.3** se refiere a evaluar la calidad de los algoritmos implementados en las Act. 1 y 2. Se realizarán mediciones de desempeño utilizando diversas evaluaciones de speed –up (Bertsekas y Tsitsiklis, 1997) y de fracciones seriales (Karp y Flatt, 1990).

En función de las conclusiones generadas en la Act. 3, se elaborarán otras propuestas de diseño (**Act. 4**) que complementen el trabajo ya realizado. Una alternativa podría ser dividir el problema de diseño en dos partes: la primera en la que se determinen las cantidades requeridas de cada familia de compuestos y luego, mediante el uso del algoritmo genético, determinar qué compuesto de cada familia es más representativo de la misma. De esta forma se disminuye el número de variables del problema y por lo tanto, el tiempo de resolución. Esta tarea es posible debido a que se tiene un amplio conocimiento del comportamiento de equilibrio de fases de los compuestos presentes en la gasolina. Otra alternativa podría ser el desarrollo de algoritmos basados en dos metaheurísticas de trayectoria: Simulated Annealing (Kirkpatrick y Vecchi, 1983) y Tabu Search (Glover, 1994). Esto daría lugar a nuevas

variantes de resolución del problema de optimización. Inspirado en Tabu Search, sería interesante incluir una lista en la que se almacenen los individuos que ya fueron analizados de forma tal de no volver a invertir tiempo en los mismos.

En el transcurso de todas estas actividades, el alumno redactará informes escritos semanales para documentar el trabajo realizado. Además, con el objeto de agilizar el desarrollo, se efectuarán reuniones grupales con la participación activa del alumno para discutir los progresos y establecer los nuevos rumbos que convendrá tomar.

Se proyecta llevar a cabo estas actividades según el siguiente cronograma:

Actividad	**Mes 1**	**Mes 2**	**Mes 3**	**Mes 4**
1	X			
2		X		
3	X	X	X	
4			X	X

Referencias

Bertsekas D.P., Tsitsiklis J.N. Parallel and Distributed Computation: Numerical Methods, Athena Scientific, (1997).

Cantú-Paz, E. A Survey of Parallel Genetic Algorithms., Calculateurs Paralleles, Reseaux et SystPmes ripartis 10 (1998) 141-170.

González Prieto M., Sanchez F.A., Pereda S., Thermodynamic modeling of new generation biofuels, X Iberoamerican Conference on Phase Equilibria and Fluid Properties for Process Design, EQUIFASE (2015a), Alicante, España.

González Prieto M., Sanchez F.A., Pereda S., Thermodynamic model for product and process design in Biorefineries, 11th International Conference on Renewable Resources and Biorefineries, RRB-11 (2015b), York, United Kingdom.

Glover F., Tabu Search Fundamentals and uses, University of Colorado at Boulder., (1994).

S. Jonuzaj, P.T. Akula, P.-M. Kleniati, C.S. Adjiman, The formulation of optimal mixtures with generalized disjunctive programming: A solvent design case study, AIChE J. (2015).

Karp A.H., Flatt H.P., Measuring Parallel Processor Performance, Comm. ACM 33 (1990) 539-543.

A.T. Karunanithi, L.E.K. Achenie, R. Gani, A New Decomposition-Based Computer-Aided Molecular/Mixture Design Methodology for the Design of Optimal Solvents and Solvent Mixtures, Ind. Eng. Chem. Res. 44 (2005) 4785–4797.

Kirkpatrick S., Vecchi M. P., Optimization by simulated annealing, Science 220 (1983), 671-680.

J. A. Klein, D.T. Wu, R. Gani, Computer aided mixture design with specified property constraints, Comput. Chem. Eng. 16 (1992) S229–S236.

N.A. Yunus, K. V. Gernaey, J.M. Woodley, R. Gani, A systematic methodology for design of tailor-made blended products, Comput. Chem. Eng. 66 (2014) 201–213.