

Optimización multicriterio por análisis envolvente de datos:

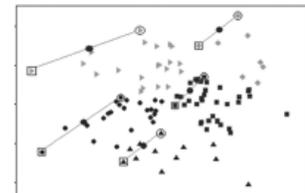
Estrategias de agrupamiento y discriminación

Matilde Luz Sánchez Peña, M. Guadalupe Villarreal Marroquín,
Mauricio Cabrera Ríos

Programa de Posgrado en Ingeniería de Sistemas, FIME-UANL
matitasanchez@gmail.com, lupita.villarreal@gmail.com,
mcabrera@mail.uanl.mx

RESUMEN

El Análisis Envolvente de Datos ha sido utilizado exitosamente en la manufactura para resolver problemas de optimización multicriterio. Su planteamiento, sin embargo, requiere de gran cantidad de datos y de software especializado de optimización. Esto impacta negativamente el costo de este tipo de proyectos. El optimizador disponible en MS Excel ha sido utilizado en estos casos analizando todos los datos por partes. Aunque factible, su utilización requiere resolver una gran cantidad de problemas de optimización. Se ha utilizado el agrupamiento de datos por métodos estadísticos para reducir el número de problemas de optimización, pero también se deterioró la fidelidad de la frontera eficiente encontrada. En este trabajo se intenta mejorar esta debilidad seleccionando mejores puntos representativos por grupo y se exploran algunas ideas para discriminar datos para reducir el tamaño inicial de los problemas.



PALABRAS CLAVES

Optimización multicriterio, agrupamiento de datos, análisis envolvente de datos.

ABSTRACT

Data Envelopment Analysis has been successfully used in manufacturing to solve multiple criteria optimization problems. However, its use implies the analysis of large amounts of data and specialized optimization software. The latter negatively impacts the cost of this kind of projects. MS Excel's optimizer has been explored to be used in these cases, by analyzing the data in a subgroup-by-subgroup basis. Although it is a feasible idea, the number of optimization problems to be solved is quite large. Data clustering through statistical techniques has been tried, resulting in considerably fewer optimization problems to be solved but with a loss of the fidelity of the efficient frontier. In this work, an improvement of the fidelity of the efficient frontier is sought after through the selection of better representative points per group and some ideas on data discrimination to make the problems smaller are explored.

KEYWORDS

Multiple criteria optimization, data clustering, data envelopment analysis.

INTRODUCCIÓN

La optimización en el terreno de la manufactura se ha caracterizado por el uso de técnicas con una sola función objetivo que representa una medida de desempeño (MD). Sin embargo, en la realidad la mayoría de los problemas en esta disciplina involucran múltiples MDs. Un mayor inconveniente resulta del hecho que estas MDs se encuentren en conflicto. Por ejemplo, se puede encontrar en muchos procesos de manufactura un conflicto entre la calidad final de un producto y su tiempo de ciclo.

El problema matemático que recoge los elementos de un problema con múltiples MDs en conflicto es el problema de optimización de criterios múltiples. La resolución de problemas de este tipo, al contrario de los que comprenden solamente una MD, no busca una solución óptima, sino un conjunto de soluciones llamadas soluciones eficientes o Pareto-eficientes. Estas soluciones definen una frontera eficiente, sobre la cual se encuentran los mejores compromisos alcanzables para ambas MDs.

La figura 1 muestra una representación gráfica para el problema mencionado. En ella se representan soluciones candidatas en términos de sus valores en las MDs de interés: tiempo de ciclo en segundos que se desea minimizar y calidad, representada por un índice y que se desea maximizar. Dada la descripción de los objetivos, una solución ideal estaría en la esquina superior izquierda de la gráfica, sin embargo no existe una solución ahí. Lo que se puede encontrar en el conjunto de soluciones candidatas son las soluciones en la frontera del conjunto o soluciones eficientes. Las soluciones eficientes están unidas por líneas en la figura 1 para mostrar la frontera eficiente. Como se puede notar, las soluciones eficientes nunca pueden mejorar en todas las medidas de desempeño al mismo tiempo, mientras que las no eficientes sí. La conveniencia de una representación gráfica se pierde al aumentar el número de MD's en el problema y es por eso que se tiene que recurrir a técnicas matemáticas para resolver este tipo de problemas.

La discretización que se presenta en la figura 1, propuesta y utilizada por Cabrera-Ríos, *et al.*¹⁻⁸ permite la aplicación de la técnica no paramétrica basada en programación lineal conocida como Análisis Envolvente de Datos (AED) para encontrar la frontera eficiente. La técnica de AED fue

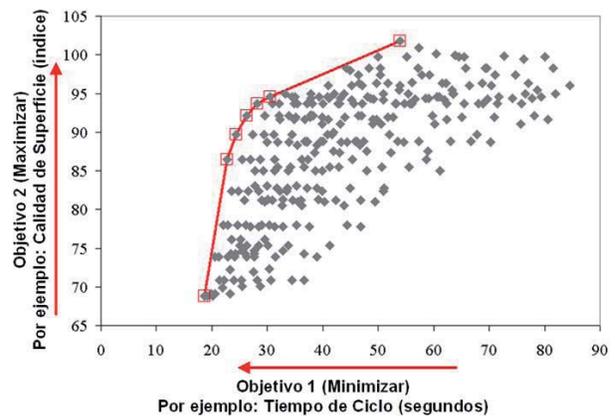


Fig. 1. Representación del problema de optimización de criterios múltiples. Las soluciones candidatas señaladas por rombos sólidos, se encuentran dominadas por una serie de puntos eficientes -unidos por líneas- que forman una frontera eficiente. Este problema en particular involucra dos criterios u objetivos: minimizar el tiempo de ciclo y maximizar la calidad de la superficie.

desarrollada por Charnes, Cooper y Rhodes.^{9,10} Aunque se ha reportado éxito en la utilización del AED en la solución de problemas de optimización de criterios múltiples, uno de los retos para llevar este método a la práctica tiene que ver con el número de soluciones candidatas que se tienen que evaluar. Un número alto de soluciones afecta negativamente: 1) la cantidad de veces que el modelo de AED se tiene que resolver y 2) el tamaño del problema que se tiene que resolver. Tener un alto número de soluciones candidatas, sin embargo, es necesario para que la fidelidad de la frontera eficiente que se encuentre sea mayor.

Algunos ejemplos reales del uso de AED para resolución de problemas de criterios múltiples llegan a tener más de 10,000 soluciones candidatas.^{2,4,5,6,8} Esto dificulta la instauración del método en la industria, pues optimizadores económicos como el MS Excel Solver, sólo pueden resolver problemas pequeños. En particular, el MS Excel Solver estándar sólo puede dar cabida a un máximo de 200 restricciones. La adquisición de software más potente afectaría negativamente el costo de los proyectos de optimización, por lo que lograr una metodología basada en el optimizador mencionado redundaría en el beneficio de una transferencia tecnológica a la industria.

Los resultados de Villarreal y Cabrera-Ríos¹¹ apuntan a que se puede eficientar la solución de los problemas de criterios múltiples a través de AED en situaciones reales de la industria por medio de la aplicación de métodos de agrupamiento. En dicho trabajo se comprobó empíricamente que el uso del agrupamiento estadístico de datos reduce significativamente el tiempo computacional en comparación con utilizar los datos sin agrupar. Sin embargo, al utilizar las intersecciones de las medianas de los grupos (el punto donde se intersecta la mediana de cada MD) como medida representativa, se sacrificó la cantidad de puntos eficientes reales encontrados. En este trabajo se exploran estrategias que permitan mantener la reducción de tiempo de análisis lograda anteriormente, pero aumentando el número de puntos eficientes encontrados. Estas estrategias incluyen la elección de puntos representativos diferentes a la intersección de las medianas. Por otro lado, se exploran también estrategias de discriminación de datos para reducir las dimensiones del problema multicriterio original.

EXPERIMENTACIÓN Y RESULTADOS

A continuación se presentan las metodologías utilizadas, clasificadas en dos divisiones: (1) agrupamiento de datos y selección de puntos representativos y (2) discriminación de datos.

En los análisis se hizo uso de la herramienta construida por Villarreal y Cabrera-Ríos.¹¹ Esta herramienta es una instauración de AED en MS Excel capaz de evaluar 100 vectores a la vez y se utiliza para determinar los puntos eficientes de este conjunto.

Para los estudios que se describen en la siguiente sección, se utilizaron 10,000 datos iniciales (vectores en R^2) los cuales se obtuvieron por medio del generador de números aleatorios en Excel. Los datos provienen de distribuciones normales con media 500 y desviación estándar 20.

Metodologías con agrupamiento de datos y selección de puntos representativos

En este primer estudio, se partió de un agrupamiento por el método de las k-medias, el cual se encuentra disponible en el paquete computacional Minitab. Este método separa los datos disponibles

en k grupos minimizando la distancia euclidiana alrededor de las k medias. Los detalles se pueden consultar en.¹² Tomando los 10,000 datos originales se obtuvieron 100 grupos (con diferente número de vectores cada uno). La figura 2 ilustra un ejemplo reducido del agrupamiento por este medio: cinco grupos denotados por diferentes figuras. Dado este agrupamiento se seleccionó un punto en los grupos como dato representativo como se describe a continuación:

I. Intersección de Medianas de cada grupo como medida descriptiva

Este método fue utilizado por Villarreal y Cabrera-Ríos, sin embargo, al reducir el tiempo, también se reducía la cantidad de puntos eficientes encontrados exitosamente. Aquí se vuelve a presentar este método para compararlo contra los desarrollos del trabajo presente. Los pasos son como sigue:

- 1) Se calculan las medianas de cada grupo en sus dos MDs y se toma la intersección de éstas como punto representativo de cada uno de los 100 grupos. En la figura 2 se ilustran estos datos representativos para nuestro ejemplo señalados con cuadrados vacíos.
- 2) Se analizan los datos con la herramienta de AED y se obtienen los grupos eficientes.
- 3) Los grupos eficientes son divididos en sus componentes.
- 4) El total de vectores componentes es analizado con la herramienta de AED para identificar la frontera eficiente. Cuando los componentes son más de 100, se volverán a crear 100 grupos utilizando k-medias y se iterará desde el paso 1.

II. Medias de cada grupo como medida descriptiva

El procedimiento es similar al anterior sólo que en lugar de las medianas se utiliza la intersección de las medias de cada grupo en sus MDs (representadas con círculos vacíos en la figura 2).

III. El punto con el máximo valor en la MD a maximizar

En esta variante se toma el punto con el máximo valor en la MD a maximizar como punto representativo de cada grupo y se sigue el mismo

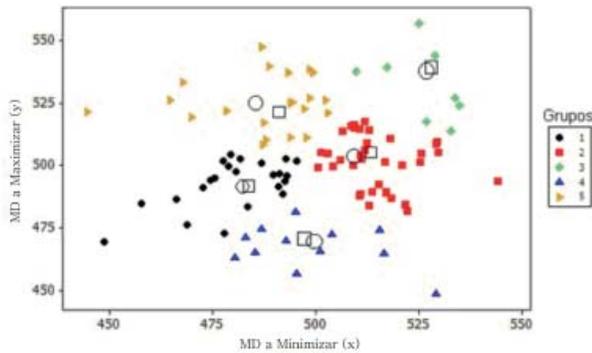


Fig. 2. Ilustración del cálculo de las medianas (cuadrados vacíos) y medias (círculos vacíos) por cada grupo.

procedimiento que en el caso base involucrando las medianas. En la figura 3, estos puntos representativos se señalan con un círculo vacío y la MD a maximizar se representa con el eje y.

IV. El punto con el mínimo valor en la MD a minimizar

Este es un caso similar al anterior, pero se toma el valor mínimo en la MD a minimizar como punto representativo. En la figura 3, los puntos de este caso se representan con cuadrados vacíos y la MD se asocia al eje x.

V. Combinación lineal convexa del punto con el máximo valor en la MD a maximizar y el punto con el mínimo valor en la MD a minimizar

Utilizando los puntos encontrados en los dos casos anteriores, es posible derivar un punto que sea

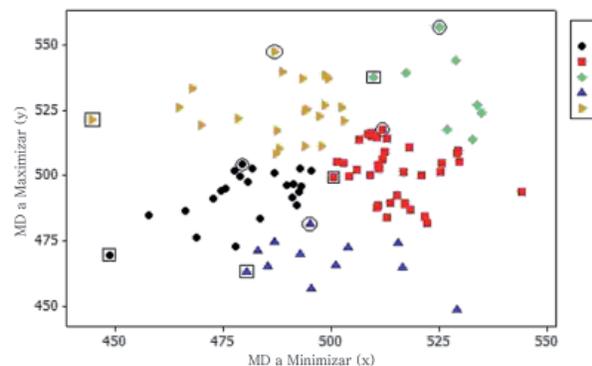


Fig. 3. Selección de los valores mayores en y (círculos) y valores menores en x (cuadrados) como datos representativos en el agrupamiento de datos.

la combinación lineal convexa de ambos y utilizarlo como punto representativo por grupo. La figura 4 ilustra más claramente los puntos creados; éstos se encuentran en el punto medio de la recta que une a los datos utilizados en los dos métodos anteriores.

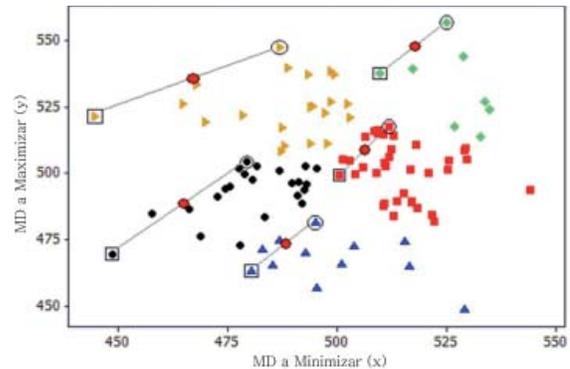


Fig. 4. Representación gráfica del cálculo de la combinación convexa de los datos máximos en y y mínimos en x.

Los resultados de los cinco casos con estrategias de agrupamiento se presentan en la tabla I.

Tabla I. Resultados comparativos. Se muestran los puntos eficientes encontrados, el porcentaje que éstos representan del total y la cantidad de ejecuciones del AED para cada metodología propuesta.

Metodología	Cantidad de puntos eficientes encontrados en proporción a los puntos eficientes originales	Cantidad de evaluaciones de AED
Intersección de medianas de los grupos	3/4	208
Intersección de medias de los grupos	3/4	208
Puntos de cada grupo con el máximo valor en y	4/4	171
Puntos de cada grupo con el mínimo valor en x	4/4	200
Combinación convexa de puntos con máximo valor en y y puntos con mínimo valor en x	4/4	220

En la tabla I se ilustra la cantidad de puntos eficientes obtenidos por cada método en proporción de la cantidad de puntos totales pertenecientes a la frontera eficiente, así como la cantidad de ejecuciones del AED para llegar a ésta.

Los métodos que utilizaron medidas estadísticas como la mediana y la media fueron menos competitivos en la calidad de la frontera eficiente, como ya había acusado la primera parte de este trabajo.¹¹

Utilizando puntos representativos diferentes se obtuvo una mejor calidad y, en dos de los casos, incluso una ligera disminución de evaluaciones con AED.

DISCRIMINACIÓN DE DATOS

En esta sección se estudian dos métodos de discriminación de datos que no aportan información de valor para llegar a la frontera eficiente. Estos métodos, a diferencia de los anteriores, se enfocan a reducir el número de puntos a analizar en un problema multicriterio en el contexto aquí descrito.

I. DISCRIMINACIÓN POR EXTREMOS

Partiendo de los datos disponibles en los que se busca minimizar una MD y maximizar otra, la posición ideal para estos fines se encuentra en la esquina superior izquierda. Si se genera una recta de 45° en dicha esquina y se desplaza hacia la esquina opuesta, se tendrá contenida la totalidad de la frontera eficiente del lado superior izquierdo en un punto dado. Ésto haría innecesario el tomar en cuenta los valores completamente en uno de los lados de la diagonal para encontrar la frontera eficiente (figura 5).

Con base en esta idea, se generó un esquema en el que el punto de referencia es la intersección de los puntos medios de los rangos de los datos, denominado en este trabajo “punto medio”. Adicionalmente, en preparación del método se realizan dos ordenamientos: uno de mayor a menor con respecto de la MD a maximizar y otro de menor a mayor en la MD a minimizar.

Tomando como referencia el punto medio (\bar{x}, \bar{y}) se puede definir que el área donde $x \leq \bar{x}$ abarca los cuadrantes 1 y 2 en la figura 6 y que el

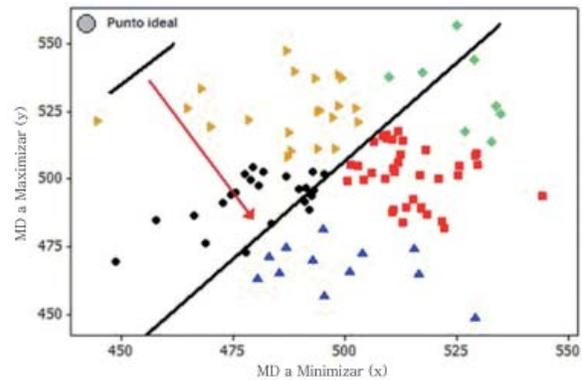


Fig. 5. Descripción gráfica del planteamiento para la discriminación por extremos.

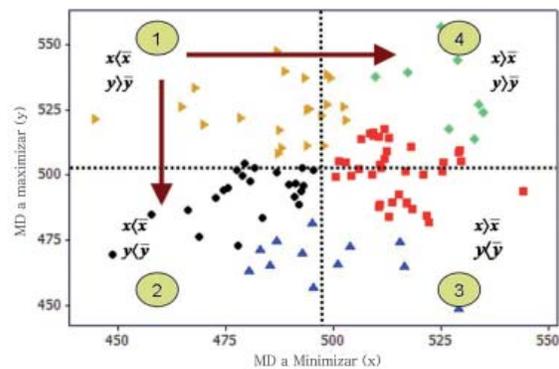


Fig. 6. División en cuadrantes tomando como criterio el punto medio absoluto de nuestros datos, las flechas indican el sentido en que los datos fueron ordenados en función de las diferentes MDs.

área donde $x > \bar{x}$, los cuadrantes 3 y 4. Por otro lado, considerando y , sabemos que los cuadrantes 1 y 4 contendrán los puntos que sean $y > \bar{y}$ y los cuadrantes 2 y 3 los puntos que sean $y \leq \bar{y}$

En este ejemplo, la frontera eficiente se encontrará dentro de los cuadrantes 1, 2 y 4, dejando los puntos del cuadrante 3 dominados. Considerando los ordenamientos mencionados, se tendrá que en el ordenamiento de mayor a menor en y aparecerán primero los datos de los cuadrantes 1 y 4 y al final los de los cuadrantes 2 y 3. Del mismo modo, con respecto a x aparecerán primero los datos de los cuadrantes 1 y 2, dejando al final los elementos del grupo 3 y 4.

A partir de los ordenamientos se toman los datos extremos como representativos en el total de nuestros datos tal como se señala en la figura 7.

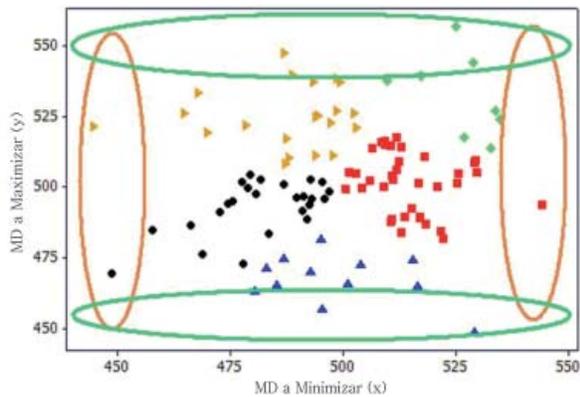


Fig. 7. Selección de datos representativos para la discriminación por extremos.

En este caso con apoyo gráfico, se tomó una cantidad de datos que contendrían la frontera eficiente de los extremos del ordenamiento, quedando los pasos de la siguiente manera:

- 1) Se realizan dos ordenamientos, uno decreciente en función de la MD a maximizar, y el otro creciente en función de la MD a minimizar.
- 2) Se seleccionan los 100 datos iniciales y los 100 finales de cada ordenamiento.
- 3) Se analizan estos 400 datos en la herramienta de AED.
- 4) Los resultantes como puntos eficientes, se vuelven a analizar para comprobar su pertenencia a la frontera eficiente.

Con este método se reduciría grandemente la cantidad de puntos a tratar por medio del AED; sin embargo, no es fácil determinar dónde se debe ubicar la diagonal para que contenga completamente la frontera eficiente, especialmente cuando más de dos MDs estén involucradas en el problema.

Otra limitación de este método es que la cantidad seleccionada de datos es determinada arbitrariamente. La definición sistemática y competitiva de este número de datos a tomar, así como pruebas adicionales con este método se dejan como trabajo a futuro.

II. Discriminación por punto Nadir

En un problema de optimización de criterios múltiples, existen varios puntos de referencia. Dos puntos importantes de este tipo son el punto nadir y el punto ideal.

Dados el punto con máximo valor en la MD a maximizar asociada al eje y y con coordenadas (x_{max} , y_{max}) y el punto con el mínimo valor en la MD a minimizar (eje x), con coordenadas (x_{min} , y_{min}). El punto ideal se encuentra en las coordenadas (x_{min} , y_{max}). Por otro lado, el punto nadir tendrá coordenadas (x_{max} , y_{min}). Estos puntos se representan en la figura 8.

Entonces, en dos dimensiones podemos hallar el punto nadir para eliminar los valores que caigan fuera del rectángulo definido por los puntos nadir e ideal. La figura 8 presenta esquemáticamente esta situación.

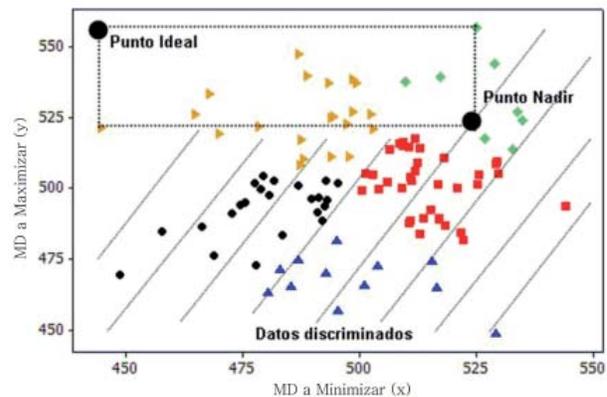


Fig. 8. Descripción gráfica de la localización del punto ideal y su correspondiente punto Nadir para la discriminación de datos.

Los pasos quedan como sigue:

- 1) Se encuentran los puntos ideal y nadir
- 2) Se eliminan los puntos por fuera de la zona definida por los puntos ideal y nadir.

En casos de dos dimensiones, encontrar el punto nadir es fácil, así que por ahora sirve a los propósitos de este estudio, sin embargo, una vez que se aborde la extensión de estos métodos a un mayor número de dimensiones, se hará más complicado encontrar el punto nadir. Esto constituye por ahora la debilidad de este método.

La evaluación de estas técnicas de reducción de puntos, iniciando con el total de 10,000 puntos, se presenta en la tabla II. También se presenta en esta tabla la cantidad de soluciones eficientes contenidas en los puntos que no fueron discriminados. Como se puede ver, aunque a distinta escala, las reducciones

Tabla II. Resultados comparativos. Se muestra el número de puntos que quedaron después de aplicar los métodos de discriminación descritos, así como las soluciones eficientes contenidas en los puntos filtrados.

Método	Número de puntos después de discriminación	Número de soluciones eficientes después de discriminación
Discriminación por extremos	400 (de 10,000)	4/4
Discriminación por punto Nadir	6,426 (de 10,000)	4/4

son significativas. Por otro lado, es interesante ver que se pudo mantener intacto el número de soluciones eficientes originales.

Si las dificultades técnicas que se explicaron en los métodos de discriminación de datos se pueden resolver convenientemente en el futuro, la combinación de estas técnicas con las de agrupamiento estadístico podría hacer aún más atractivo la utilización de AED para problemas reales en la industria.

CONCLUSIONES

En este trabajo se exploraron métodos que utilizan diferentes puntos representativos a partir del agrupamiento estadístico de datos para aumentar la fidelidad de la frontera eficiente en la resolución de problemas de optimización de múltiples criterios con Análisis Envolvente de Datos. Se exploraron además algunas ideas iniciales en discriminación de datos.

Los resultados presentados apuntan a que se puede reducir significativamente el tiempo de análisis por medio de estrategias de agrupamiento de datos y, tras una selección cuidadosa de puntos representativos, mantener un nivel aceptable de fidelidad en la frontera eficiente final. También se demostró el potencial de reducir el tiempo computacional por medio de una discriminación inicial de datos. Los métodos presentados, sin embargo, deberán ser mejor estudiados en el futuro.

AGRADECIMIENTOS

La realización de este proyecto fue posible gracias al apoyo UANL-PAICYT CA-1069-05 y a PROMEP 103.5/04/2590. Se reconoce también el valioso apoyo CONACYT, la FIME y la UANL en términos de las becas de las estudiantes involucradas en el desarrollo de este proyecto.

REFERENCIAS

1. M. Cabrera-Ríos y J.M. Castro, The Balance Between Durability, Reliability, and Affordability in Structural Composites Manufacturing: Preliminary Results, Technical Paper 2003-01-0459, Reliability and Robust Design in Automotive Engineering, 2003 SAE World Congress Book SP-1736, March 2003.
2. M. Cabrera-Ríos, Thesis (PhD), www.ohiolink.edu/etd/, The Ohio State University (2002).
3. M. Cabrera-Ríos, M., K.S. Zuyev, X. Chen, J.M. Castro, and E.J. Straus, Optimizing Injection Gate Location and Cycle Time for the In-Mold Coating (IMC) Process, *Polymer Composites* 23:5 (2002).
4. C. E. Castro, M. Cabrera-Ríos, B. Lilly, J.M. Castro, and C.A. Mount-Campbell, Identifying The Best Compromises Between Multiple Performance Measures In Injection Molding (IM) Using Data Envelopment Analysis (DEA), *Journal of Integrated Design & Process Science*, 7, 77-87 (2003).
5. M. Cabrera-Ríos, J.M. Castro, and C.A. Mount-Campbell, Multiple Quality Criteria Optimization In In-Mold Coating (IMC) With A Data Envelopment Analysis Approach, *Journal of Polymer Engineering*, 22:5 (2002).
6. M. Cabrera-Ríos, J.M. Castro, y C.A. Mount-Campbell, Multiple Quality Criteria Optimization In Reactive In-Mold Coating With A Data Envelopment Analysis Approach II: A Case With More Than Three Performance Measures, *Journal of Polymer Engineering*, 24:4 (2004) 435.

7. M. Castro, M. Cabrera-Ríos, y C.A. Mount-Campbell, Modelling and Simulation in Reactive Polymer Processing, Modelling and Simulation in Materials Science and Engineering, 12 (2004) S121-S149.
8. C.E. Castro, M. Cabrera-Ríos, B. Lilly, and J.M. Castro, Simultaneous Optimization of Mold Design and Processing Conditions in Injection Molding, Aceptado para publicación en Journal of Polymer Engineering (2005).
9. A. Charnes, W.W. Cooper, and E. Rhodes, European Journal of Operational Research, 2:6 (1978).
10. A. Charnes, W.W. Cooper, A.Y. Lewin, and L.M. Seiford, Data Envelopment Analysis: Theory, Methodology, and Applications. Boston. Kluwer Academic Publishers (1993).
11. M.G. Villarreal y M. Cabrera-Ríos, Agrupamiento de Datos para la Solución del Problema de Optimización Multicriterio. Ciencia UANL, 10:2, 137-142 (2007).
12. J. Han and M. Lamber, Data Mining Concepts and Techniques, Morgan Kaufmann Publishers, 335-391 (2001).
13. M. A. Urbano-Vázquez y M. Cabrera-Ríos, Comparación de diseños experimentales para la predicción de líneas de costura en el moldeo por inyección: Resultados Preliminares. Memorias Congreso CIINDET (2006).

Anúnciese en:

Ingenierías



INFORMES:
Tel: (52) (81) 83294020 Ext. 5854
Fax: (52) (81) 83320904
E-mail: revistaingenierias@gmail.mx
Internet: <http://ingenierias.uanl.mx>