



## Agrupamiento de datos para la solución del problema de optimización multicriterio

MA. GUADALUPE VILLARREAL MARROQUÍN\*, MAURICIO CABRERA RÍOS\*

La optimización en el terreno de la manufactura se ha caracterizado, principalmente, por el uso de técnicas con una sola función-objetivo que representa una medida de desempeño (MD). En realidad, la mayor parte de los problemas en esta disciplina involucran múltiples MDs. Más aún, estas MDs están frecuentemente en conflicto. Un ejemplo es la manufactura de partes con materiales poliméricos, en la que se involucran interacciones entre diversos fenómenos físicos y químicos, generalmente acoplados por las diversas temperaturas de los procesos de producción. La variación de una de estas temperaturas afecta varios fenómenos simultáneamente y no a todos de una manera deseable.

La consideración de sólo una MD para su optimización puede resultar en combinaciones de variables de decisión que determinen niveles muy pobres en otras MDs, dependiendo del nivel de conflicto entre éstas. De ahí que la consideración simultánea de las MDs de interés sea indispensable en presencia de conflicto.

El problema matemático que recoge los elementos de un problema con múltiples MDs en conflicto es el problema de optimización de criterios múltiples. Frecuentemente, en los tratados, a este problema se le llama programación de objetivos

múltiples, nombre que toma a raíz de las funciones objetivo que representan a las MDs. La resolución de tales problemas, al contrario de aquéllos donde solamente se considera una MD, no busca una solución única, sino un abanico de soluciones conocidas como *Pareto-eficientes* o simplemente *soluciones eficientes*. El conjunto de estas soluciones define una *frontera eficiente*. Una representación gráfica conveniente de este problema se encuentra en la figura 1. En esta figura se grafican soluciones candidatas en términos de sus valores en las MDs de interés: tiempo de ciclo en segundos y calidad de superficie representada por un índice. Se requiere minimizar el tiempo de ciclo y maximizar la calidad de superficie de una parte fabricada con resina poliéster reforzada como objetivos. Una solución ideal estaría en la esquina noroeste de la gráfica (lo más cercano a (0,105)); sin embargo, no hay una solución ahí. Las soluciones eficientes se encuentran en la frontera de los datos creada por las direcciones deseables de ambos objetivos. Las soluciones eficientes, unidas por líneas a manera de mostrar la frontera eficiente en la figura 1, nunca pueden mejorar en todas las medidas de desempe-

\*División de Posgrado en Ingeniería de Sistemas, FIME-UANL.

ño al mismo tiempo, mientras que las no eficientes sí. La conveniencia de una representación gráfica se pierde al aumentar el número de MDs en el problema, y por eso se tiene que recurrir a técnicas matemáticas para resolver los problemas resultantes.

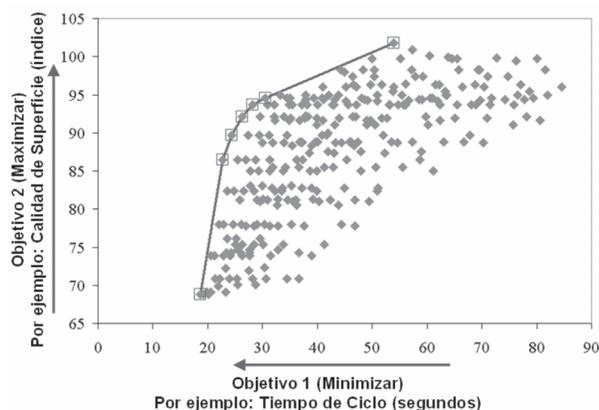


Fig. 1. Representación del problema de optimización de criterios múltiples. Las soluciones candidatas, señaladas por rombos, se encuentran dominadas por una serie de puntos eficientes unidos por líneas, formando una frontera eficiente. Este problema en particular involucra dos criterios u objetivos: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

La discretización de la región factible que se presenta en la figura 1, propuesta y utilizada por Cabrera Ríos et al.<sup>1,8</sup> permite la aplicación de la técnica no paramétrica conocida como Análisis envolvente de datos (AED), la cual se basa en la programación lineal, para encontrar la frontera eficiente. La técnica fue desarrollada por Charnes, Cooper y Rhodes<sup>9,10</sup> (se remite al lector a las referencias 4-10 para conocer más a fondo acerca de ella). Aunque se ha reportado éxito en la utilización de AED, 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. Este número de soluciones afecta negativamente: 1) la cantidad de veces que el modelo de AED se tiene que resolver; 2) el tamaño del problema que se tiene que resolver. Tener un alto número de soluciones candidatas, sin embargo, es necesario para que la calidad de las soluciones eficientes que se encuentren sea más alta.

Algunos ejemplos reales del uso de AED para resolución de problemas de criterios múltiples lle-

gan a tener más de 10,000 soluciones candidatas.<sup>2,4,6,8</sup> Esto implica tener que resolver el problema de programación lineal del AED de un tamaño de más de 10,000 restricciones, más de 10,000 veces. Esto dificulta la instauración del método en la industria, pues el optimizador comercial más económico del mercado, MS Excel Solver, sólo puede resolver problemas con 200 restricciones. Aunque una solución a este caso sería adquirir software más potente, esto impactaría negativamente el costo de un proyecto de optimización y, muchas veces, lo hace perder su atractivo ante la industria regional y nacional.

## Planteamiento del problema

Este proyecto se enfocó en el estudio de metodologías de agrupamiento de datos para hacer más eficiente el uso del AED al resolver problemas reales de optimización de criterios múltiples. Con este objetivo se incluyó el uso de métodos estadísticos y redes neuronales artificiales (RNAs) para agrupar puntos en espacios  $n$ -dimensionales. Esta última técnica de agrupamiento proviene de la idea de varios investigadores, dentro de los que se distinguen a Grossberg<sup>11</sup> y a Coñeen,<sup>12-14</sup> en términos de mapas autoorganizados y de aprendizaje competitivo.

El componente básico de este trabajo consistió en la integración efectiva de técnicas de áreas distintas (optimización e inteligencia artificial) para determinar un esquema de aplicación tecnológica viable para la industria. La hipótesis fue que se podrían resolver de manera económica los problemas de criterios múltiples en situaciones reales de la industria a través de la combinación adecuada de métodos que puedan aportar sus fortalezas a distintas fases de la solución de estos problemas.

A continuación se describe los componentes del estudio:

**Datos.** Los datos con los que se contó provinieron de simulaciones numéricas en las que se cuenta con múltiples MDs en conflicto.

**Variables.** Las variables de decisión en este trabajo se asocian a las diferentes técnicas de agrupación: utilización de métodos estadísticos o redes neuronales artificiales.

**Restricciones.** Dado el objetivo de este trabajo,

las técnicas que se utilizaron debían resultar en un esquema para aplicar el AED en MS Excel, esto es, como una consecución eficiente que contuviera solamente hasta 200 restricciones.

*Objetivo.* Encontrar la combinación de técnicas y parámetros de agrupamiento que ayuden al AED a converger a la frontera eficiente, minimizando el número de corridas, respetando la capacidad del Solver de Excel.

Un esquema de agrupación exitoso no sólo contribuirá a hacer una transferencia tecnológica económica a la industria, sino que hará factible discretizar los problemas de criterios múltiples a niveles arbitrarios (lo cual haría tener una mayor fidelidad de la frontera eficiente), confiando en que se pueden resolver de manera efectiva y conveniente.

## Aplicación y resultados

En esta sección se presentan las técnicas utilizadas para el agrupamiento de datos y los resultados de la aplicación.

Primeramente se construyó una herramienta de AED instaurada en MS Excel, la cual evalúa 100 vectores a la vez y puede determinar los vectores *eficientes* de este conjunto. Se consideraron 10,000 datos iniciales (vectores en  $R^2$ ), los cuales se obtuvieron con el generador de números aleatorios incluido en MS Excel, con éstos se realizaron los siguientes tres estudios utilizando diferentes técnicas de agrupamiento de datos:

### 1. A modo de torneo

En esta primera prueba para encontrar la frontera eficiente se realizó una competencia a modo de torneo y, por lo tanto, una evaluación exhaustiva de todos los datos de la siguiente manera:

- 1) Se extrajeron, de los 10,000 vectores, 100 grupos de 100 vectores cada uno, los cuales se escogieron en el orden en que se encuentran los datos. Los primeros 100 vectores corresponden al primer grupo y así sucesivamente.
- 2) Dados estos  $n$  grupos ( $n \in [1,100]$ ), se introdujeron los 100 vectores de cada uno a la herramienta de AED, con la cual se determinaron los vectores *eficientes* de cada grupo. Esto resultó en

100 evaluaciones (una por cada 100 vectores).

- 3) Si el número de vectores *eficientes* resultantes del *paso 2* era menor que 100, se introducían éstos nuevamente a la herramienta de AED y se determinaban los vectores *eficientes*, terminando aquí el análisis.
- 4) Si no sucedía lo anterior, se extraían nuevamente grupos de 100 vectores (puede haber un grupo con menos de 100 vectores) y se regresaba al *paso 2*.

### 2. Utilizando redes neuronales artificiales

En esta prueba se utilizó una red neuronal artificial (RNA) *competitiva* para agrupar los 10,000 vectores en 100 grupos (en ocasiones puede converger a un número menor de grupos, como sucedió en este estudio). La RNA fue creada con ayuda del paquete matemático Matlab™. Para este caso se realizó lo siguiente:

- 1) Dados los  $n$  grupos ( $n \in [1,100]$ ), se calculó la *mediana* de cada uno, de lo cual resulta un vector representativo por cada grupo.
- 2) Los vectores resultantes del *paso 1* se introdujeron a la herramienta de AED, de éstos los vectores resultantes en la frontera eficiente determinan los grupos *eficientes*.
- 3) Los grupos *eficientes* del *paso 2* se separaron en sus componentes.
- 4) Si el número de vectores resultantes en el *paso 3* era menor que 100, se introducían directamente a la herramienta de AED, para determinar los vectores *eficientes* y se finalizaba la prueba.
- 5) Si no, se reagrupaban nuevamente en 100 grupos utilizando RNAs y se regresaba al *paso 1*.

### 3. Utilizando métodos de agrupamiento estadístico

En este caso se utilizó el método estadístico de agrupamiento de  $k$ -medias para formar 100 grupos de datos. El método se encuentra disponible en el paquete computacional estadístico Minitab™, y los detalles técnicos se pueden consultar en J. Han y M Lamber.<sup>15</sup> Con estos grupos se realizó lo siguiente:

- 1) Primero se calculó la *mediana* de cada conjunto

- de vectores, para obtener un representante por grupo, esto es, obtener 100 vectores.
- 2) Se introdujeron estos 100 vectores a la herramienta de AED para determinar los grupos eficientes.
  - 3) Los grupos eficientes se dividieron en sus componentes.
  - 4) Si el número de vectores de estos grupos era menor a 100, se introducían a la herramienta de AED y se obtenían los vectores eficientes, con los cuales se terminaba la prueba.
  - 5) Si resultaban más de 100 vectores, se reagrupaban en 100 grupos nuevamente con los métodos de *k*-medias y se regresaba al *paso 1*.

Los resultados de los tres casos descritos anteriormente se presentan en la tabla I, la cual está organizada de la siguiente manera: en la primera fila se describen las diferentes medidas de evaluación; la segunda fila contiene los resultados de utilizar exhaustivamente AED con MS Excel; en la tercera se presentan los resultados del esquema de agrupamiento a través de RNAs, y en la cuarta los resultados de agrupar datos mediante métodos estadísticos. En la primera columna se muestra la metodología utilizada, seguida del número de vectores originales, el número de rondas de análisis que se realizaron (cada vez que se realizan los pasos presentados anteriormente para cada método), el número de evaluaciones (análisis de 100 vectores en la herramienta de AED), el tiempo estimado tanto de evaluación como agrupamiento, el tiempo total de análisis (tiempo total estimado por agrupamiento más el tiempo total estimado de evaluación). Finalmente, se presenta el porcentaje de puntos eficientes obtenidos que en realidad son puntos eficientes del problema original.

En la figura 2 se muestran dos de las medidas más destacadas de la tabla I: el tiempo total estimado de análisis y el porcentaje de puntos (vectores) eficientes correctos (número de vectores eficientes encontrados entre el número de vectores eficientes reales).

En la figura 2 se observa que al utilizar métodos estadísticos para agrupamiento se obtuvo el beneficio de una significativa reducción de tiempo de análisis con el costo de perder confiabilidad. Usar

RNAs para agrupamiento fue excesivamente costoso en tiempo, además de pérdida de confiabilidad. El uso exhaustivo de AED a través de MS Excel resultó con un alto tiempo de análisis, aunque menor al de la RNA y, por supuesto, con la certeza de encontrar todas las soluciones eficientes del problema.

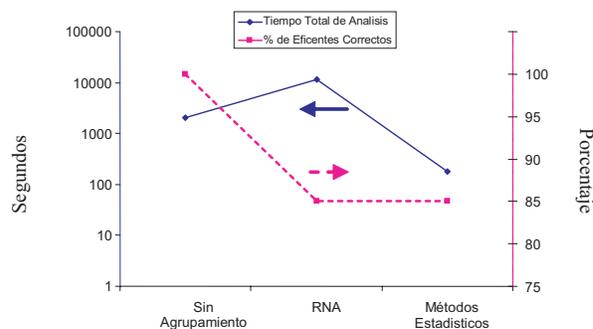


Fig. 2. Tiempo total estimado de análisis (línea sólida y eje de la izquierda) y porcentaje de puntos eficientes (línea punteada y eje de la derecha).

## Conclusiones y trabajo a futuro

En este trabajo se presentaron los resultados de utilizar diferentes metodologías de agrupamiento de datos para hacer más eficiente el uso del Análisis envolvente de datos, al solucionar problemas de optimización de múltiples criterios.

Se demostró empíricamente que es factible reducir significativamente el tiempo de análisis aplicando los métodos propuestos para determinar una frontera eficiente. Ante la pérdida de confiabilidad al agrupar soluciones por redes neuronales artificiales y métodos estadísticos, se vislumbra que utilizar el AED de manera exhaustiva utilizando paquetería computacional modesta (MS Excel) puede ser la manera más segura de aplicación en la industria. Sin embargo, si el tiempo de análisis es muy importante, el agrupamiento con técnicas estadísticas resulta muy atractivo.

De lo anterior quedan varios puntos abiertos para trabajos posteriores, dentro de los que se en-

Tabla I. Resultados comparativos. \*Dado que en el caso de agrupamiento mediante RNAs fue demasiado costoso (tiempo) realizar el agrupamiento, se descartó este método como accesible para agrupar en rondas subsecuentes. En este caso se utilizó la técnica del caso 1 después de la primera ronda.

Herramienta	Puntos originales	Rondas	Número de evaluaciones	Número de agrupamientos	Tiempo estimado por evaluación (seg)	Tiempo estimado por agrupamiento(seg)	Tiempo total estimado de análisis(seg)	Número de puntos eficientes	Porcentaje de puntos eficientes correctos
A modo de torneo	10,000	3	104	0	20	0	2,080	7	100%
RNA	10,000	3	27	1*	20	10,800	11,340	7	85%
Agrupamiento estadístico	10,000	3	3	2	20	60	180	7	85%

cuentran: 1) realizar una instauración de los métodos estadísticos de agrupamiento en MS Excel para no contar con la limitante de paquetería computacional especializada; 2) probar si alguna otra medida puede representar cada uno de los grupos a analizar (en este estudio se tomó la mediana); 3) dado que en los casos en los cuales no se obtuvieron todos los puntos eficientes reales fue debido a que en alguna ronda del estudio se descartó el grupo al que pertenecía, se plantea que se puedan tomar no sólo los grupos eficientes, sino además algunos no eficientes dentro de cierto rango de discriminación; 4) aplicar los métodos propuestos a un problema real.

## Resumen

En la industria de la manufactura la gran mayoría de las veces se buscan satisfacer objetivos múltiples que, en general, muestran diferentes grados de conflicto entre ellos. La importancia de la solución efectiva de este tipo de problemas es evidente. Por esta razón, en este trabajo se presenta el estudio de diferentes esquemas para el agrupamiento de datos, que hacen más eficiente el uso de Análisis envolvente de datos (AED) para resolver problemas reales de optimización de criterios múltiples. Los resultados de este estudio apuntan favorablemente a la adopción de los esquemas de agrupamiento en aplicaciones reales.

**Palabras clave:** Optimización multicriterio, Análisis envolvente de datos, Técnicas de agrupamiento de datos.

## Abstract

In the manufacturing industry it is not uncommon to find optimization problems requiring the compromise between multiple performance measures. Finding solutions in a convenient manner to this kind of problems is critical if optimization is to be applied in these cases. In this work, data grouping schemes are proposed which make Data Envelopment Analysis more efficient in solving multicriteria optimization problems. The results of this study provide support for the adoption of these grouping schemes in real applications.

ing multicriteria optimization problems. The results of this study provide support for the adoption of these grouping schemes in real applications.

**Keywords:** Multiple criteria optimization, Data Envelopment Analysis, Data clustering techniques.

## Agradecimientos

Este trabajo fue posible gracias al apoyo de la UANL por medio del proyecto Paicyt CA1069-05. Los autores agradecen además las becas Conacyt para los estudiantes asociados a 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. (2002). Multiple Criteria Optimization Studies in Reactive In-Mold Coating, Thesis (PhD), www.ohiolink.edu/etd/, The Ohio State University.
3. M. Cabrera-Ríos, K.S. Zuyev, X. Chen, J.M. Castro, E.J. Straus. (2002). Optimizing Injection Gate Location and Cycle Time for the In-Mold Coating (IMC) Process, Polymer Composites 23:5.

4. C. E. Castro, M. Cabrera-Ríos, B. Lilly, J.M. Castro, C.A. Mount-Campbell. (2003). 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.
5. M. Cabrera-Ríos, J.M. Castro, C.A. Mount-Campbell. (2002). Multiple Quality Criteria Optimization in In-Mold Coating (IMC) With A Data Envelopment Analysis Approach, *Journal of Polymer Engineering*, 22:5.
6. M. Cabrera-Ríos, J.M. Castro, C.A. Mount-Campbell. (2004). 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*, 24:4 435.
7. M. Castro, M. Cabrera-Ríos, C.A. Mount-Campbell. (2004). Modelling and Simulation in Reactive Polymer Processing, *Modelling and Simulation in Materials Science and Engineering*, 12 S121-S149.
8. C.E. Castro, M. Cabrera-Ríos, B. Lilly, J.M. Castro. (2005). Simultaneous Optimization of Mold Design and Processing Conditions in Injection Molding, Aceptado para publicación en *Journal of Polymer Engineering*.
9. A. Charnes, W.W. Cooper, E. Rhodes. (1978). Measuring the Efficiency of Decision Making Units, *European Journal of Operational Research*, 2:6.
10. A. Charnes, W.W. Cooper, A.Y. Lewin, L.M. Seiford. (1993). *Data Envelopment Analysis: Theory, Methodology, and Applications*. Boston. Kluwer Academic Publishers.
11. S. Grossberg, *Nonlinear Neural Networks: Principles*. (1988). *Mechanisms and Architectures, Neural Networks*, 1, 17-61.
12. T. Kohonen. (1982). Self-Organized Formation of Topologically Correct Feature Maps, *Biological Cybernetics*, 43, 59-69.
13. T. Kohonen. (1984). *Self-Organisation and Associative Memory*, Springer-Verlag, Berli.
14. T. Kohonen. (1990). The Self-Organizing Map, *Proceedings of the IEEE*, 78, 1464-1480.
15. J. Han and M Kamber. (2001). *Data Mining Concepts and Techniques*, Morgan Kaufmann Publishers, 335-391.

*Recibido: 16 de junio de 2006*

*Aceptado: 6 de marzo de 2007*