

Impacto del agrupamiento de escenarios en un modelo de optimización no determinista de planificación de la producción y distribución

Enrique Fernández¹, Javier Marengo²

¹ Instituto de Industria, Universidad Nacional de General Sarmiento,
Universidad Católica Argentina
efernandez@campus.ungs.edu.ar, enrique_fernandez@uca.edu.ar,
<https://orcid.org/0009-0009-4413-8371>

² Escuela de Negocios, Universidad Torcuato Di Tella,
javier.marengo@utdt.edu,
<https://orcid.org/0000-0003-2694-4758>

The impact of scenario clustering on a non-deterministic optimization model for production and distribution planning

Resumen

En el presente trabajo se analizan dos técnicas clásicas de agrupamiento sobre un conjunto de escenarios que fueron utilizados en un problema optimización de la producción y distribución con características no determinista que fueron generados mediante un procedimiento que combina el método de Montecarlo con un proceso de difusión. Se analizan los resultados del utilizando escenarios representativos de cada grupo generado con la realización de la optimización utilizando los escenarios sin agrupar.

Palabras claves: agrupamiento, optimización, simulación

Abstract

In this paper, two classic clustering techniques are analyzed on a set of scenarios that were used in a production and distribution optimization problem with non-deterministic characteristics. These scenarios were generated using a procedure that combines the Monte Carlo method with a diffusion process. The results are analyzed using representative scenarios from each group generated with the optimization performed using the ungrouped scenarios.

Keywords: clustering, optimization, simulation

1. Introducción

Se presenta el problema de producción y distribución de un producto que se utiliza para neutralizar el efecto que produce un agente sobre un cultivo.

Se plantea un modelo multiperíodo de optimización lineal entera mixta estocástica para planificar la producción necesaria y el procedimiento de distribución.

Un plan óptimo de producción y distribución depende de la demanda del producto, la cual es, esencialmente no determinista. Se simula esta demanda sobre la base de un mecanismo de generación de escenarios, cada uno de los cuales tiene asociada una probabilidad de ocurrencia.

Cómo se ha presentado en oportunidades anteriores ([2],[3],[4]) la cantidad de variables involucradas en el modelo determina que no se encuentre solución exacta en un tiempo de cómputo aceptable, por lo cual se diseñaron un conjunto de heurísticas que ajustan la solución óptima de la relajación lineal, buscando resultados cercanos al óptimo en un tiempo de cómputo aceptable.

Con la intención de incrementar los escenarios que participan en la solución del modelo en el presente **trabajo en curso**, se exploran técnicas de clustering para obtener grupos que puedan contar con un escenario representativo de cada conjunto. Se exploran alternativas para la creación del escenario representativo de un conjunto y se comparan los resultados de las soluciones encontradas utilizando los escenarios que representan a los grupos contra los resultados obtenidos con los escenarios sin agrupar.

2. Técnicas de clasificación

Las técnicas de clasificación permite agrupar objetos por características similares las cuales pueden ser supervisadas o no supervisadas [1].

En las técnicas supervisadas se etiqueta cada uno de los elementos con una o más etiquetas y luego se realiza el agrupamiento de aquellos elementos que tienen etiquetas similares, en cambio el agrupamiento no supervisado trabaja con elementos no etiquetados buscando similitudes entre ellos a partir de los valores de las variables

por las cuales se desean realizar los grupos.

Se encuentran referencias a varios algoritmos de clasificación supervisada basadas en el uso de redes neuronales [6] como la utilización del método de enjambre de partículas (PSO) [10], etc.

La clasificación no supervisada, usualmente llamada clustering cuenta con diversidad de técnicas que se pueden agrupar en las jerárquicas y las basadas en partición.

Los métodos de clustering jerárquicos conforman los subconjuntos dividiendo iterativamente el conjunto original mediante un patrón descendente o uniendo los subconjuntos en un patrón ascendente.

Los métodos jerárquicos descendentes van dividiendo el conjunto original en subconjuntos hasta que se llega a conformar subconjuntos individuales o hasta que se cumpla un condición de terminación, en cambio los ascendentes parte de los elementos individuales que se van agrupando en conjuntos hasta conformar uno solo o hasta que se cumplan ciertas condiciones de terminación. [9]

3. Características de la experimentación numérica

En el presente trabajo se exploran varios métodos clásicos de agrupamiento para generar grupos de escenarios que fueron determinados con anterioridad en un procedimiento del tipo Montecarlo asociado a un mecanismo de difusión descrito en [3] que participan en un procedimiento de optimización.

Inicialmente se toman dos métodos clásicos como k-means [1] y single-linkage [5] para crear los grupos a partir de los cuales se construye el escenario representativo de cada uno de ellos.

En ambos casos se utiliza la misma función de distancia:

$$d_{i,t} = \frac{\sum_{s=1}^n |x_{t,s} - x_{i,s}|}{n} \quad (1)$$

La ecuación 1 obtiene el valor $d_{i,t}$ que indica la distancia del elemento i del conjunto a agrupar al centroide del grupo t que se obtiene a partir de $x_{t,s}$ que indica el valor que toma el atributo de clasificación s en el centroide del grupo t , $x_{i,s}$ que es el valor del atributo de clasificación s del elemento i del conjunto a agrupar y n la cantidad de variables en cada elemento del conjunto de escenarios a agrupar.

A continuación, se utiliza la heurística de dos fases presentada en [2] utilizando por un lado, los escenarios representativos y por el otro los escenarios sin agrupar. Se comparan los resultados de ambas ejecuciones.

Como se indica en [8] la función $f(k)$ utilizada para evaluar el resultado de usar k conjuntos debe contemplar varios factores:

- la función de distancia utilizada (sesgo de proximidad),
- el nivel de detalle requerido por el observador,
- la interdependencia entre los diferentes conjuntos.

Se evaluará la cantidad de grupos a conformar en ambos métodos mediante la función

$$f(k) = \frac{1}{N} \sum_{t=1}^k \sum_{j=1}^{n_t} d_{j,t} \quad (2)$$

Esta ecuación 2 obtiene una medida de la distancia promedio de los elementos del conjunto a agrupar al centroide del grupo al cual fue asignado dado que $d_{j,t}$ indica la distancia del elemento j al centroide t asociado al conjunto al cual pertenece j .

Cuando $\frac{f(k)}{f(k-1)}$ es un valor cercano a 1 indica que no se obtienen mejoras sustanciales al incrementar la cantidad de grupos en los cuales se separa el conjunto. Se busca este valor y se contrasta con la relación sugerida en [7] entre la cantidad de elementos del conjunto a agrupar y la cantidad de grupos que se generan:

$$k \approx \sqrt{\frac{N}{2}}$$

con N la cantidad de elementos del conjunto a agrupar y k la cantidad de grupos a crear.

Se repite la experimentación con una cantidad k de grupos inferior y otra superior a la sugerida en los párrafos anteriores.

Utilizando el método single-linkage se agrupa el conjunto original en k grupos comparando el valor de la función $f(k)$ con la obtenida mediante el método de k-means y los tiempos de cómputo requeridos en ambos agrupamientos.

Para terminar se evalúa las soluciones encontradas mediante el agrupamiento contra las soluciones de la heurística de dos fases descrita en [2] con los escenarios sin agrupar.

Referencias

1. Amit, S., Mukesh, P., Akshansh, G., Neha, B., Om Prakash, P., Aruna, T., Meng, J. E., Weiping, D., Chin-Teng, L. (2017). A review of clustering techniques and developments, *Neurocomputing* (267), 664-681.
2. Fernández, E., Marengo, J. (2022). Algoritmos Heurísticos para un modelo de optimización estocástica de planificación de la producción y distribución - Heuristic algorithms for a stochastic optimization problem in production and distribution planning, XXI Latin Ibero – American conference on Operations Research.
3. Fernández, E., Marengo, J. (2019). Optimización Lineal Robusta de un problema de fabricación / distribución con demanda estocástica sujeta a un proceso de difusión. Estudio de un caso de distribución de un producto agrícola para el cultivo de soja – Estudio comparativo de tres heurísticas, Simposio de Investigación Operativa en el marco de las 48 Jornadas Argentinas de Informática e Investigación Operativa.
4. Fernández, E., Marengo, J. (2018). Optimización Lineal Robusta de un problema de fabricación / distribución con demanda estocástica sujeta a un proceso de difusión. Estudio de un caso de distribución de un producto agrícola para el cultivo de soja – Abordaje mediante heurísticas, Simposio de Investigación Operativa en el marco de las 47 Jornadas Argentinas de Informática e Investigación Operativa.
5. Jain, A. K., Murty, M.N., Flynn, P.J. (1999). Data Clustering: A Review, *ACM Computing Surveys* 31(3).
6. Kröse, B., Van der Smagt, P. (2018). An introduction to Neural Networks. The University of Amsterdam, 3, 23-27.
7. Madhulatha, T. S. (2012). An overview on clustering methods. *IOSR Journal of Engineering*, 2(4), 719-725.
8. Pham, D. T., Dimov, S. S., Nguyen, D. D. (2005). Selection of K in K-means clustering. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*. 219(1), 103-119. (2005)
9. Xu, D., Tian, Y., A Comprehensive Survey of Clustering Algorithms. *IEEE Transactions in Cybernetics*, 2, 165-193.
10. Xue, B., Zhang, M., Browne, W. (2013). Particle swarm optimization for feature selection in classification: A multi-objective approach. *IEEE Transactions in Cybernetics*, 6, 1656-1671.