

SISTEMAS DE RECOMENDACIÓN PARA DATOS EN ALTA DIMENSIÓN: UNA NUEVA PROPUESTA METODOLÓGICA BASADA EN CESTAS DE CONSUMO

Maria Florencia Acosta
FICH-UNL, Argentina
ma.flor.acosta@gmail.com

Los sistemas de recomendación son herramientas matemáticas que, a partir de datos nos recomiendan productos o servicios. Los más conocidos son los que utilizan las plataformas de transmisión de contenido (streaming), pero cada vez se utilizan más en comercio electrónico, bancos, plataformas de enseñanza, entre otros.

Un sistema de recomendación no es más que un método de filtrado que toma la información relevante para el problema, descartando la información que no es completamente informativa para el mismo. La mayoría de los métodos de recomendación se basan en factorización de matrices, y pueden ser del tipo colaborativo o no colaborativo. El primero, se basa en utilizar la información de usuarios para realizar la recomendación, mientras que el segundo solo utiliza la información del usuario en cuestión. Estos métodos son basados en datos (data-driven) y por lo tanto son métodos automáticos, que necesitan ser entrenados a partir de bases de datos confiables.

Para el caso particular de métodos de recomendación basados en análisis de cestas de consumo, la cantidad de productos involucrados en el problema puede ser significativamente mayor a la cantidad de cestas, por lo que el problema se torna de alta dimensionalidad, surgiendo en este caso una matriz de cesta-productos dispersa (sparse). Las metodologías clásicas utilizadas en este tipo de problemas generalmente utilizan matrices de cesta-productos binarias, reglas de asociación y/o medidas de similaridad que no contemplan la alta dimensionalidad del problema.

En el presente trabajo proponemos un nuevo método basado en aglomerado (clustering) que utiliza una matriz de cesta-productos sparse compuesta por la participación en las ventas totales de cada producto, donde las recomendaciones surgen de acuerdo a la similaridad de las cestas de consumo pero considerando el peso que tiene cada producto en las ventas totales. A su vez, se utiliza una medida de similaridad apta para alta dimensionalidad de los datos, buscando pesar los agrupamientos con otros factores relevantes para el sistema de recomendación como ser el tamaño del cliente, la asignación del gasto, y la importancia del ítem recomendado en los ingresos por ventas. Mas aún, este método resulta invariante ante cambios generalizados de precios, resultando así adecuado en contextos inflacionarios.

La motivación de esta metodología surge de la necesidad de una firma mayorista que vende alrededor de 1500 productos alimenticios y busca recomendar productos a sus clientes considerando no sólo la probabilidad de compra sino también su relevancia al ingreso por venta generado.

Trabajo en conjunto con Rodrigo García Arancibia (UNL & CONICET), Pamela Llop (FIQ-UNL & CONICET) y Mariel Guadalupe Lovatto (FIQ-UNL & CONICET).

Referencias

- [1] Sarkar, Soham and Ghosh, Anil K, On perfect clustering of high dimension, low sample size data, IEEE transactions on pattern analysis and machine intelligence, volume 42, number 9, pages 2257-2272 , 2019, IEEE.
- [2] Hahsler, Michael and Grün, Bettina and Hornik, Kurt, Arules-A computational environment for mining association rules and frequent item sets, Journal of statistical software, volume 14, number 15, pages 1-25, 2005, University of California at Los Angeles.
- [3] Boztg, Yasemin and Reutterer, Thomas, A combined approach for segment-specific market basket analysis, European Journal of Operational Research, volume 187, number 1, pages 294-312, 2008, Elsevier.
- [4] Reutterer, Thomas and Dan, Daniel, Cluster analysis in marketing research, Handbook of market research, 221-249, 2021, Springer.

[5] Hahsler, Michael and Karpienko, Radoslaw, Visualizing association rules in hierarchical groups, Journal of Business Economics, volume 87, pages 317-335, 2017, Springer.