

Índice general

1. Análisis de cesta de la compra (Market basket análisis)	3
1.1. Técnica	4
1.2. Indicadores de las reglas de asociación	4
1.3. Algoritmo a priori	5
Bibliografía	5

Capítulo 1

Análisis de cesta de la compra (Market basket análisis)

En este capítulo revisaremos el método de análisis de cesta de la compra con el fin de identificar patrones o asociaciones entre diversos grupos de productos. Esta metodología pretende, además de la validación de distintas asociaciones que se pueden considerar obvias debido a una reiterada compra conjunta, encontrar relaciones entre productos cuya asociación no es tan evidente. La principal fuente de información para llevar a cabo este tipo de análisis serán datos que recogen las transacciones de los clientes en cada compra.

El concepto de reglas de asociación fué conocido en el año 1993 tras la publicación de un artículo de Agrawal, R; Imielinski, T.; Swami, A. *Mining association rules between sets of items in large databases*.

A continuación vamos a definir una serie de conceptos para entender mejor este método. Una *asociación* es una concurrencia de dos o más cosas. Por ejemplo, los perritos calientes pueden estar asociados con los refrescos, los cebolla o el ketchup. Se dice que existe una asociación *positiva* si la presencia de algunos productos implica la presencia de otros dentro de la misma transacción. También existe la asociación *negativa*, la cual consiste en que la presencia de algunos productos haga muy improbable la presencia de otros elementos en la misma transacción.

Se conoce como *reglas de asociación* a la agrupación de productos en función de la afinidad existente entre ellos. La utilidad de estas reglas se encuentra en la identificación de oportunidades y el diseño de diversos grupos de productos que puedan ser atractivos para los consumidores. Estas reglas tienen como finalidad el descubrimiento de las relaciones implícitas en los datos. Cada una de estas reglas consta de un antecedente y un consecuente. Por ejemplo, en la regla siguiente regla de asociación: Si un consumidor compra un perrito caliente, también tiende a comprar un refresco, cebolla y ketchup. Aquí el perrito caliente es el antecedente, y el refresco, la cebolla y el ketchup son los consecuentes.

Cuando aplicamos un análisis de cesta de la compra a unos datos, habitualmente podemos obtener tres posibles resultados generados a partir de las reglas de asociación:

- **Resultados factibles:** Contienen información útil y de calidad, ya que los patrones identificados son factibles. Estas reglas nos podrían ayudar en la toma decisiones como por ejemplo, saber a que productos se aplicar una promoción con el objetivo de impulsar las ventas de los mismos y de sus productos relacionados.

- **Resultados triviales:** Estos son conocidos por aquellos que están familiarizados con el negocio en cuestión. Este tipo de reglas nos muestra productos en los que no se plantea la compra de un producto sin otro, por ejemplo, la compra de pintura requiere la necesidad de comprar también pinceles.
- **Resultados imposibles:** Estos resultados implican la obtención de reglas incongruentes, ya que nos proporciona una información que no facilita el entender el comportamiento del consumidor, sino que se trata de casos puntuales en un momento determinado.

1.1. Técnica

Podemos expresar las reglas de asociación de la siguiente forma: $A \rightarrow B$, donde A e B son conjuntos de ítems. Sea $I = \{i_1, i_2, \dots, i_m\}$ ítems y sea D una serie de transacciones, siendo T una transacción. Se tiene que $T \subseteq I$, y por tanto, el conjunto I tendrá que ser redefinido como binario $\{0, 1\}$. La secuencia de valores de cada transacción T , será generada a partir de la identificación de atributos de I con valor 1.

Se conoce como itemset A , al conjunto de artículos $A \subset I$, por tanto, cada transacción T contiene al itemset A , $A \subseteq T$. Por tanto, una regla de asociación es una implicación de la forma:

$$A \rightarrow B, A \subset I, B \subset I, A \cap B = \emptyset$$

1.2. Indicadores de las reglas de asociación

Una vez obtenidas las reglas de asociación, es necesario evaluarlas. Dada la siguiente regla de asociación $A \rightarrow B$, podemos definir tres indicadores: lift, soporte y confianza. La importancia de la obtención de éstos radica en la información no redundante que aportan sobre el conjunto de reglas obtenidas

1. **Lift:** Este parámetro indica si realmente existe una asociación entre uno o varios productos y de que tipo es, si positiva o negativa. Su valor viene dado por:

$$lift = \frac{P(B|A)}{P(B)P(A)}$$

Es decir, la probabilidad de comprar dos ítems a la vez (asociación) entre la probabilidad de comprarlos por separado. Por eso, podemos decir que este parámetro indica la fuerza de asociación entre los productos de la izquierda y de la derecha de la regla, y cuanto mayor sea su valor, mayor será el vínculo entre los productos.

Este parámetro puede ofrecernos tres posibles resultados:

- Si $lift > 1$, la probabilidad de que ambos productos estén asociados es mayor que la probabilidad de que no lo estén, considerando que éstos guardan una relación positiva.
- Si $lift \approx 1$, este resultado sugiere que los productos son independientes y su presencia conjunta se debe al azar.

- Si $lift < 1$, la probabilidad de que los productos no estén asociados es mayor que la probabilidad de que estén, indicando que los productos no están asociados y guardan una relación negativa.
2. **Soporte:** El soporte del elemento A del conjunto de transacciones D viene dado por:

$$supp(A, B) = P(B \cap A) = \frac{\text{Nº transacciones con A y B}}{\text{Nº total de transacciones}}$$

Es decir, se trata de la proporción de transacciones que contienen una combinación particular de elementos en relación con el número total de transacciones. Es un valor que está acotado: $0 \leq supp \leq 1$, y cuánto mayor sea su valor, mayor probabilidad de que la asociación considerada se repita en futuras transacciones.

3. **Confianza:** este indicador nos ofrece información sobre la precisión de la asociación, se trata de la probabilidad de comprar el elemento B, dado que el elemento A aparece en la transacción y se puede calcular de la siguiente forma:

$$C(A \rightarrow B) = P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{\text{Nº transacciones con A y B}}{\text{Nº de transacciones en las que aparece A}}$$

Al igual que el soporte, su valor está acotado: $0 \leq C \leq 1$.

1.3. Algoritmo a priori