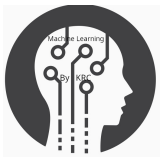
Machine Learning (Association Rule Mining)

Prof K R Chowdhary

CSE Dept., MBM University

January 02, 2025



Market basket (Itemset)

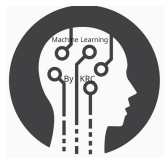
⇒ Defn.: *Market basket*. It is a well defined business activity: a collection of items purchased by a customer in an individual transaction.

⇒ Such transaction is due to customer's purchase from a grocery store, or online purchase.

⇒ Over a time, retailers accumulate huge collection of transactions of performing a business activity. An analysis performed on collections of such transactions' is: find sets of items, or *itemsets*, that appear together in many transactions.

⇒ A pattern to be extracted through this analysis consists of *itemset* and the corresponding number of transactions that contains it. Objective: use knowledge of these patterns to improve placement of items in a store, or web pages on a website, or use this to motivate potential customers who can buy these itemsets.

⇒ Task of *association rule mining* (ARM) is finding correlation between items in a dataset [1].



Basic theory

$\Rightarrow$ Association rules: set of significant correlations, frequent patterns, associations, in data.

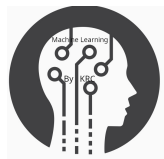
$\Rightarrow$ These are transactional databases, and repositories. Mining of association rules: capturing correlations, patterns, rules, and representing them as *if*. . . *then* rules.

$\Rightarrow$ Defn: *Association Rule*: Given a set of transactions, each comprising a set of items, an association rule is implication, $X \Rightarrow Y$, here X, Y are itemsets, indicating presence of X implies Y.

$\Rightarrow$ E.g., an insurance company finds a correlation between two policies X, Y, i.e., $X \Rightarrow Y$: customers holding policy X were also likely to hold policy Y.

$\Rightarrow$ Now, it could target marketing policy Y through those clients who hold policy X but not Y.

$\Rightarrow$ They have been applied in: *market basket analysis* in commercial environments, crime prevention, countering terrorism: a relationship between objects can be concluded as useful knowledge.



Association rule mining (ARM)

Three ways to measure association:

$\Rightarrow$ "Support." Given a set of data items, association rule has *support* s for some set of items X if s percent of transactions include all the items of set X.

$\Rightarrow$ In sales transactions, e.g., if we find that sale of some items have a significant impact on total profits, we consider that proportion as *support* threshold.

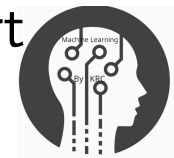
$\Rightarrow$ "Confidence:" How likely itemset Y is purchased when itemset X is purchased (find $X \rightarrow Y$). For given data items, the rule has *confidence* c if $c\%$ of transactions that contain X, also contain Y.

$\Rightarrow$ "Confidence:" Proportion of transactions with item X, in which Y also appears. For total transactions T; c is,

$$c = \frac{(X,Y)}{X} = \frac{(X,Y)/T}{X/T}$$

$$= \frac{support(X,Y)}{support(X)}$$

ARM Goal: Discover all association rules having support and confidence greater than some minimum threshold.



Association rule mining ...

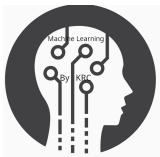
$\Rightarrow$ Def.*Lift.* The "lift" (l) says, how likely the itemset Y is purchased when itemset X is purchased, while controlling "lift" for how popular the item Y is. Lift is measured as:

$$l = \frac{support(X, Y)}{support(X) \times support(Y)}. \quad (1)$$

A lift value greater than 1 means item Y is likely to be bought if X is bought, while its value less than 1 means Y is unlikely to be bought if X is bought.

Given the sales transactions in Table 1, find out the:

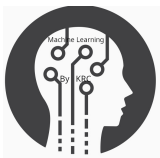
1. "Support" for following:
   1. Laptop
   2. Smartphone
   3. Laptop & Smartphone
2. "Confidence" of Music system with respect to:
   1. Laptop
   2. Smartphone
   3. Laptop & Smartphone
3. "Lift" for Laptop and Music System.



Association rule mining ...

Table 1: Transactional database

Tran. ID	Cust. ID	Item's name	Price (in $)	Date
101	201	*Laptop*	1500	8/20/2018
101	201	*Tablet*	300	8/20/2018
101	201	*Smartphone*	100	8/20/2018
102	201	*Music system*	500	8/25/2018
102	201	*Smartphone*	100	8/25/2018
103	202	*Laptop*	1500	8/30/2018
103	202	*Music system*	500	8/30/2018
103	202	*Smartphone*	100	8/30/2018



Association rule mining ..

(1). Three trans. in table, trans. nos.: 101, 102, and 103, stored in eight rows, two have Laptop, three have smartphone, and two have Laptop & smartphone combined. So, "support" for items in the same order is, 67, 100, and 67% respectively.

(2)(a). We want to compute "confidence" of "Music system" w.r.t. "Laptop":

$$c = \frac{support(laptop,\ music\ sys)}{support(laptop)}$$

$$= \frac{1/3}{2/3} = 50\%.$$

(2)(b). Similarly, "confidence" of 'Music system" w.r.t. "Smartphone" is:

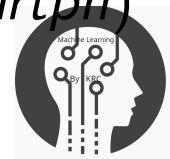
$$c = \frac{support(smartphone,\ music\ syst)}{support(smartphone)}$$

$$= \frac{2/3}{2/3} = 100\%.$$

(2)(c). Confidence of 'Music system" w.r.t. laptop & Smartphone, combined is:

$$c = \frac{support(laptop\&smartph,\ music\ s}{support(laptop\&smartph)}$$

$$= \frac{1/3}{2/3} = 50\%.$$



Association rule mining ..

(3). The "lift" for "Laptop and Music System" is computed as follows.

$$c = \frac{support(laptop,\ music\ system)}{support(laptop) \times support(music\ system)}$$
$$= \frac{1/3}{2/3 \times 2/3} = 0.75$$

"Lift" of 0.75 (< 1) indicates that item Y is unlikely to be bought by a customer who is buying item X.


Chowdhary, K.R. (2020). Data Mining. In: Fundamentals of Artificial Intelligence. Springer, New Delhi. https://doi.org/10.1007/978-81-322-3972-7_17 pp. 537-549.