IDENTIFYING CUSTOMER BUYING PATTERNS USING MARKET BASKET ANALYSIS

RAKHMANALIYEVA K.

Kazakh-British technical university, 050000, Almaty, Kazakhstan

Abstract. *Market Basket Analysis (MBA) is an approach that finds the strength of association between pairs of products that customers buy and can determine patterns of co-occurrence. The main aim of MBA is to determine customer buying behavior and predict next purchase. It can help companies to increase cross-selling.*
To generate association rules, the Apriori algorithm employs frequently purchased item-sets. It is based on the idea that a frequently purchased item's subset is also a frequently purchased item. If the support value of a frequently purchased item-set exceeds a minimum threshold, the item-set is chosen. This paper observes the advantages of implementing MBA, algorithms that applies in this technique and ways to identify customer buying patterns.

Key words: *Market Basket Analysis, Apriori algorithm, association rule, co-occurrence*

КЛИЕНТТЕРДІҢ САТЫП АЛУ ҮЛГІСІН АНЫҚТАУ ҮШІН НАРЫҚТЫҚ ҚОРЖЫНДЫ ТАЛДАУДЫ ҚОЛДАНУ

РАХМАНАЛИЕВА К.

¹Қазақстан-Британ техникалық университеті, 050000, Алматы, Қазақстан

Аңдатпа. *Нарық қоржынын талдау – бұл сатып алушылар сатып алатын тауарлардың жұптары арасындағы ассоциацияның мықтылығын айқындайтын және қатар жүру заңдылықтарын анықтайтын тәсіл. Нарық қоржынын талдаудың негізгі мақсаты клиенттердің сатып алу тәртібін анықтау және келесі сатып алуды болжау болып табылады. Бұл компанияларға кросс-сатылымды арттыруға көмектесе алады.*
Ассоциация ережелерін құру үшін, Apriori алгоритмінде жиі сатып алынатын элементтер жиынтығы қолданылады. Бұл жиі сатып алынатын заттың ішкі жиыны да жиі сатып алынатын зат деген идеяға негізделген. Егер жиі сатып алынатын заттар жиынтығының қолдау мәні ең төменгі шектен асып кетсе, элементтер жиыны таңдалады. Бұл жұмыста нарық қоржынын талдауды іске асырудың артықшылықтары, осы техникада қолданылатын алгоритмдер және клиенттердің сатып алу заңдылықтарын анықтау тәсілдері көрсетілген.

Түйінді сөздер: *нарық қоржынын талдау, апрриори алгоритмі, ассоциация ережесі*

ПРИМЕНЕНИЕ АНАЛИЗА РЫНОЧНОЙ КОРЗИНЫ ДЛЯ ОПРЕДЕЛЕНИЯ МОДЕЛИ ПОКУПОК КЛИЕНТОВ

РАХМАНАЛИЕВА К.

Казахстанско-Британский технический университет, 050000, Алматы, Казахстан

Аннотация. *Анализ рыночной корзины — это подход, который определяет силу связи между парами продуктов, которые покупают клиенты, и может определять закономерности совместного появления. Основная цель – определить покупательское поведение клиентов и спрогнозировать следующую покупку. Это может помочь компаниям увеличить объем перекрестных продаж.*
Для создания ассоциативных правил алгоритм Apriori использует часто покупаемые наборы

предметов. Он основан на идее, что подмножество часто покупаемых товаров также является часто покупаемым товаром. Если значение поддержки часто приобретаемого набора предметов превышает минимальный порог, выбирается набор предметов. В этой статье рассматриваются преимущества внедрения анализа рыночной корзины, алгоритмы, применяемые в этой методике, и способы выявления моделей покупательского поведения клиентов.

Ключевые слова: *анализ рыночной корзины, алгоритм априори, правило ассоциации.*

Introduction

Market Basket Analysis is a set of transaction data that used to identify customer attitude and determine his buying patterns. It goes by finding out the combination of things that occur together in transaction. It means that MBA lets retailers to determine relationships between things that customers purchase, find out the frequency in order to be able to predict when they will go together. These are the main benefits of MBA:

- Organizing store to increase income. Items that complement each other should be displayed together to make it easier for customers to notice them. This will determine how a store should be organized in order to maximize profits.

- Promotional message. Market basket research may increase the efficiency of any communication channel, including email, phone, social media, and direct salesperson offers. Using MBA data, you can recommend the next best product that a consumer is most likely to buy.

- Keep track of inventory. You may also estimate future purchases of clients based on MBA inputs over a period of time. You can estimate which items are likely to fall short based on your early sales data, and keep your stock in top shape.

- Content Positioning. The location of website content is critical for e-commerce enterprises. Conversions can be boosted if commodities are listed in the correct sequence. Online publishers and bloggers can utilize MBA to display content that customers are most likely to read next. This will lower bounce rates, increase engagement, and boost search results performance.

- Recommender Systems. Some well-known companies, such as Netflix, Amazon, and Facebook, already utilize recommendation algorithms. If you want to build an effective recommendation system for your business, you'll require market basket analysis to keep it running smoothly. MBA can be used as a foundation for developing a recommendation engine.

Market Basket Analysis is very helpful in evaluating qualitative data. So, it is a basic technique which big retailers like Amazon and Flipkart use to understand clients shopping habits.

Although this work implies predictive market basket analysis, there is another type of MBA - differential market basket analysis. The difference between them is that predictive MBA is the type that helps to determine cross-sell, while differential MBA considers collecting data from different stores, different client's group during different times of the day, month or year. [1] It is important to note that in this research will be used predictive type of market basket analysis.

MBA also helps to develop and expand marketing approach, like:

• Changing the store layout according to trends

• Customer behavior analysis

• Catalog design

• Cross marketing on online stores

• What are the trending items customers buy?

It is important to note that retailing it is not the only area where MBA can be used. It can be involved in banking, business, bioinformatics, manufacturing industry. So, in [2] author implemented market basket analysis technology to one of Six Sigma's phases - Improve phases. It helps to improve behavior of customer by producing association rules between products. The received association is based on general rule induction. As a result, author formed two groups of customer and generated rules for each of them. Such implementation can be also used to target special offers. In addition of applying MBA in marketing, there is another work [3]. The paper provides more detail information about association rule and describe each step of ARM-Pre-

dictor algorithm that can be helpful in determining customer behavior. Authors also point that market basket analysis gives the understanding of customers purchase behavior. The concept of ARM-Predictor algorithm is cooperating Association Rule Miner, so it helps to establish strong relationship between goods stored in shops. The basis of algorithms that find association rules is different metrics like support, confidence, lift etc.

Professors T. Raeder and N. V. Chawlathe in [4] defined the function of market basket analysis as a determining actionable knowledge in transaction databases. Which means next case: a common stock sells a big set of products P. Define a *transaction p ⊆* P as the set of items customer buys in a single trip to the store. T = {p} is the set of all transactions processed by the store in a certain time period in the transaction database of the store. However, authors took a different way to mining transaction data by representing data as a product network. So, authors try to develop the power and clarity of MBA by creating transactional data as a network. This approach can help to establish relationships between goods which is quite challenging with traditional association rules. So, authors stated that the network representation of transaction data lets using different algorithms once unavailable to the association rule community.

However, in [5] authors pointed that Apriori algorithm is the biggest improvement and easier to implement compared to other algorithms like FP-Tree Algorithm and RARM. In addition, professors Kronberger and Affenzeller [6] say that the main advantage of Apriory algorithm is the possibility of scaling big data bases with millions of items. On the other hand, researched from University of Chile [7], offer their own way of performing MBA. Their methodology produces effective and appropriate frequent item sets. It's based on graph mining techniques. Authors say that results from traditional algorithms like Apriori algorithm are meaningless since clusters were formed by enormous amount of different goods. As a result of this work were obtained 30 clusters while using k-means algorithm gave only 2 clusters.

Authors Hossain, Sarowar Sattar, Kumar

Paul used Apriori and FP Growth algorithms for market basket analysis. They proposed a new method for mining association rules that involves selecting a specific percentage of frequent items from their dataset. [8]

Methodology

The common way of identifying the relationship between items is applying set of rules called Association rules. The idea of Association rules is giving result as rules in form If This (A) Then That (B). A and B are also known as an 'antecedent' and 'consequent'.

Commonly the relationship will be in the form of a rule: IF beer, no bar meal THEN crisps. The chance that a person will buy beer without a bar meal is cited to as the support for the rule. The confidence refers to the conditional probability that a consumer will buy crisps.

The proportion of transactions that include all of the items in an itemset is known as support. The higher the support the more frequently the itemset occurs. This value allows to determine the rules worth regarding further analysis. For example, in a total of 10,000 transactions, one would want to consider only the itemsets that appear at least 50 times, i.e., support = 0.005. If support value is low, it means that we do not have enough information an itemset about relationship between items and we can't do any conclusion.

$$Support = \frac{(A+B)}{TOTAL}$$

The likelihood that a transaction containing the items on the left-hand side of the rule (beer, bar meal) also contains the item on the right-hand side is called confidence (crisps). When the confidence is high, the chance that the item on the right-hand side will be purchased is bigger. In other words, expected the return rate is getting greater.

$$Confidence = \frac{(A+B)}{A}$$

Lift is the product of the probabilities of all the items in a rule appearing together divided by the product of the possibilities of the items on the left- and right-hand sides appearing as if they had no relationship. Overall, the strength of interaction between the goods on the left and right sides

of the law is summarized by raise. When the lift is large, then link between items is greater.[9]

$$Lift = \frac{\frac{(A+B)}{A}}{\frac{B}{Total}}$$

This method is quite easy on the mathematical level. It's also an unsupervised learning tool. Applying this tool requires minimal data preparation.
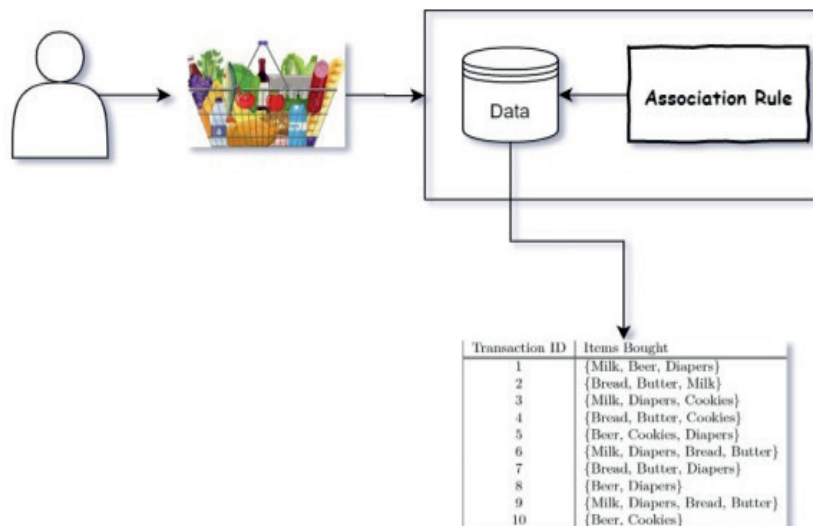

Fig. 1. Applying association rules

Apriori algorithm is useful tool in mining frequent itemset and defining association rule. This algorithm determines frequent specific items and extends them large itemset until itemsets appears more frequently in the database. Apriori is the algorithm that was first used in mining of frequent itemeset. It was proposed by Agrawal and Srikant in 1994. The alternative approach of this algorithm is joining and prune items in order to reduce looking space. The fundamental idea of Apriori algorithm is its anti-monotonicity of support measure. The implementation of Apriori algorithm includes steps illustrated in picture below [10]:
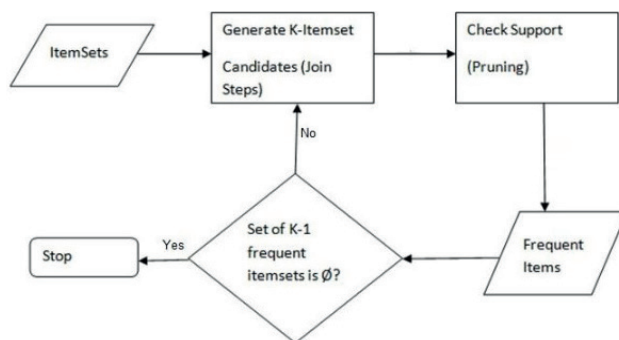

Fig. 2. Apriori algorithm steps

1) Determine the support of itemsets (of size k = 1) in database. This step is called generating the candidate set.

2) Reduce the number of candidates set by excluding items with a support less than the given threshold.

3) JRepeat the above steps until no more itemsets can be generated by joining the frequent itemsets to form sets of size k + 1. Which means that support of formed sets should be less than given support. [11]

The next step is applying Apriori algorithms on dataset in Python. The dataset was taken from UCI Machine Learning repository.[12]

It contains information about all transactions between 01/12/2010 and 09/12/2011 for online retails. The implementation in Python requires certain libraries, in this case we used mlxtend, numpy and pandas libraries. Mlxtend implements a variation of algorithms and widely used in machine learning and data mining. We cleaned the data and splitted based on area of transaction. After that we defined function to make the data suitable, so we can build model. The part of code can be seen below. Here we build model and collect the inferred rules in a dataframe. [13]

```
# Building the model
frq_items   =   apriori(basket_Hong_Kong,
min_support = 0.05, use_colnames = True)
# Collecting the inferred rules
rules = association_rules(frq_items, metric
="lift", min_threshold - 1)
 rules = rules.sort_values(['confidence', 'lift'],
ascending =[False, False])
print(rules.head())
```

In order to determine how association rule works we also conducted the survey. In this survey was applied Likert scale method. This method allows interviewees to show in what extend they agree or disagree with a particular statement. It's commonly used to get a more nuanced picture of people's attitudes and beliefs than a simple «yes/no» query. This method implies questions and 5 or 7 options to answer. So based on these answers we can measure the assumption. In this survey people were given 5 options to respond:

• strongly agree
• agree
• neutral
• disagree
• strongly disagree

A Likert scale also allows measuring other variations. For example, frequency, quality, importance, etc.

Results

Implementing Apriori algorithm for Italian transaction showed that 'WOODLAND CHARLOTTE BAG' and 'ABC TREASURE BOOK BOX' paired together. So, we can assume that instead of plastic bag people prefer to use recycled bags in order to carry staff, in this case is book. Besides, analyzing results of transaction from Britain there is a set of tea-plates. We also can assume that this set is result of local traditions.

Almost the same results we got from French transaction. Paper cups and paper and plates are bought together. The reason is also could be related to cultural features, like having family party with friends.

However, we could not observe the results from all regions. For example, while applying model for Japan there was memory error, since the data is too big.

It is also important to mention results of conducted survey. The survey was formed in order to implement Likert scale and to find out if people would buy the sets of products given in the table below together. So, we can approve or disapprove the association rate that given in Table 1. In this study were interviewed 53 people from 20 to 40 years old. Combinations were made from products that we often buy. The result of survey is shown in Fig. 3 below. Results present the percentage of people who agreed or disagreed with questions that were given in Google Form.

According to results it can be said that if people buy beer, they probably will buy cigarettes too, while the combination of products that people will not buy are milk & beer and cereals & cigarettes. It expected to prove the strong association between Milk and Bread, however, according to results we found out that association rate between these two items is 0.2. We can see that people in most cases will not buy given product together. So, it follows that expected results cannot be approved. During the survey gender, age, nationality or financial situation were not taken into account.

Table 1. The association rate between products from given data

Products	Association rate
Milk and Bread	0,9
Milk and Cereals	0,9
Milk and Beer	0,5
Milk and Chips	0,3
Milk and Cigarettes	0,2
Bread and Cereals	0,5
Beer and Chips	0,8

Chips and Cigarettes	0,4
Beer and Bread	0,2
Beer and Cigarettes	0,9

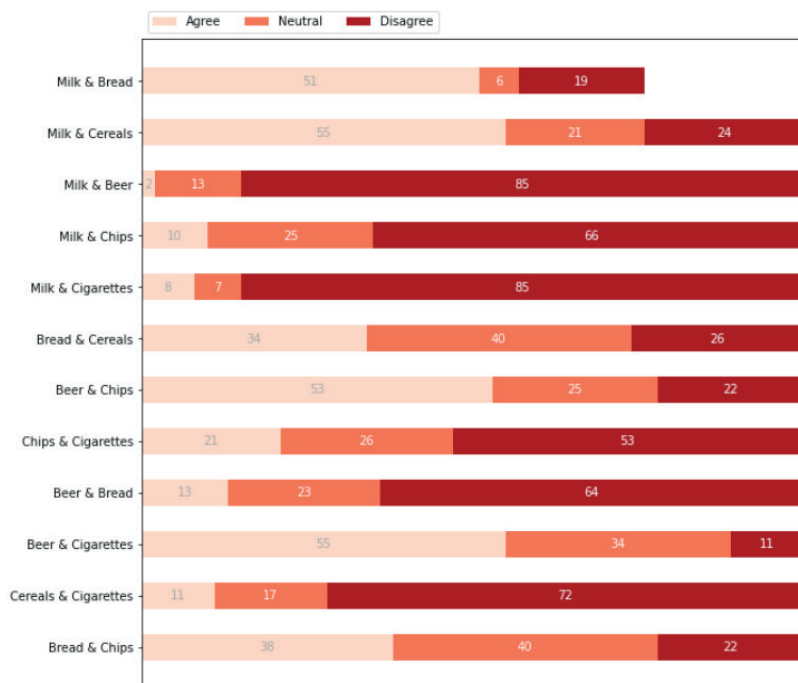| Cereals and Cigarettes	0,3
Bread and Chips	0,2

Fig. 3. Agree percentage

Conclusion

In this paper, we did a research of what the Market Basket Analysis is, where it can be used and what is the profit of applying MBA for business, did a review of previous works of researches who offers a new approach of applying algorithms in order to improve results. We tried to apply association rule and Apriory algorithm in order to identify customer buying pattern and established the associations using Python and Mlxtend library. Identifying patterns may help in

- product placement
- point-of-sale
- customer retention

We also conducted the survey in order to approve the association rate of products by applying Likert scale method. The survey showed the principles of association rules.

At this step we applied algorithm on data from open source. However, since this data is quite old and the results cannot be used by retailers nowadays, next step is trying to find real data, so we can apply the methodology on them.

The major restriction of Apriori algorithm is time, it can be slow when the data is too big, and so it makes this approach less productive. It follows to necessity using other algorithms like FP-Tree Algorithm and compare the results. As a future work we are planning to identify customer buying patterns by applying Fuzzy logic in Market Basket Analysis.

REFERENCES

1. Margaret Rouse Market Basket Analysis, https://searchcustomerexperience.techtarget.com/definition/marketbasket-analysis
2. A.Trnka *Market basket analysis with data mining methods: Six sigma methodology improvement* ICNIT 2010 - 2010 International Conference on Networking and Information Technology, 2010

3. Kaur, Manpreet Kang, Shivani *Market Basket Analysis: Identify the changing trends of market data using association rule mining* Procedia Computer Science, 2016

4. Raeder, Troy Chawla, V. Nitesh. *Market basket analysis with networks* Social Network Analysis and Mining, 2011

5. Borgelt, Christian *Frequent item set mining* Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2012

6. Kronberger, Gabriel Affenzeller, Michael *Market basket analysis of retail data: Supervised learning approach* Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2012

7. F. Ivan. Videla-Cavieres, A. R´ Sebastian ´ıos *Extending market basket analysis with graph mining techniques: A real case* Expert Systems with Applications Volume 41, Issue 4, Part 2, March 2014, Pages 1928-1936

8. Maliha Hossain, A H M Sarowar Sattar, Mahit Kumar Paul *Market basket analysis Using Apriori and FP Growth Algorithm* 22nd International Conference on Computer and Information Technology (ICCIT), 2019

9. Videla-Cavieres, F. R´Ivan ıos, A.´ Sebastian. *Extending market basket analysis with graph mining techniques: A real case* Expert Systems with Applications, 2014

10. *What is the Apriori algorithm?* https://www.educative.io/edpresso/whatis-the-apriori-algorithm

11. Lynsey McColl, *Market Basket Analysis: Understanding Customer Behaviour* https://select-statistics.co.uk/blog/market-basket-analysisunderstanding-customer-behaviour/

12. *AprioriAlgorithmInDataMining:ImplementationWithExamples* https://www.softwaretestinghelp.com/apriori-algorithm/

13. *Online Retail Data Set* https://archive.ics.uci.edu/ml/datasets/online+retail

---

Information about author:

1. Rakhmanaliyeva K. – Kazakh-British Technical University