International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 6, Issue 1, 2018

Exploring Supervised and Unsupervised Learning Techniques in Market Basket Analysis

Komal K Shah, Assistant Professor in Computer Engineering, Government Engineering College - Rajkot

Jahnvi V Doshi, Assistant Professor in Computer Engineering, Government Engineering College - Rajkot

Abstract

Market Basket Analysis is an important task in data mining. It is widely used in retail sector to reveal customer buying habits by finding associations among items in transaction data. Traditionally, unsupervised methods like Apriori or FP-Growth have been used to identify frequent itemsets and association rules in Market Basket Analysis. But with the advancement of analytical techniques, supervised learning methods have become more popular for their predictive capabilities. This paper compares supervised and unsupervised data mining methods in Market Basket Analysis. It highlights their applications, strengths, limitations and future possibilities. The paper also discusses hybrid and Fuzzy techniques to provide a wider view of modern market basket analysis.

1. Introduction

Market Basket Analysis [1] is finding out which items are commonly purchased together in customer transactions. This understanding leads to targeted marketing strategies, better store layouts, efficient inventory management and enhanced recommendation systems. Initially, Market Basket Analysis mainly focused on unsupervised learning [2] methods by mostly applying association rule mining. However the supervised learning method [3] provides a powerful method of delivering personalized insights and targeted forecasts.

As retail data grows increasingly complex including customer profiles, transaction timestamps and interactions across multiple channels, the limitations of traditional methods become clearer. This paper compares both unsupervised and supervised methods while also discusses innovative techniques like hybrid and fuzzy methods.

2. Market Basket Analysis using Unsupervised Learning

Unsupervised learning is very important in Market Basket Analysis through Association Rule Mining [2]. Unlike supervised methods that depend on labelled outcomes, unsupervised learning finds hidden patterns and relationships in transactional data without specific targets. The main goal is to discover frequent itemsets, which are groups of items that often appear together in customer transactions and to create association rules that can guide strategic business decisions. For example, a rule like "If a customer buys bread and butter, they are likely to also buy jam" shows a common buying trend. These rules are usually measured with three key metrics [4]: (i) support, which indicates how often the itemset shows up in the dataset; (ii) confidence, which tells how often the rule holds true; and (iii) lift, which measures the increased likelihood of the consequent occurring when the antecedent is present compared to random chance. These metrics are important for finding actionable insights in retail, where even minor changes in consumer behaviour can affect sales strategies.

2.1 Association Rule Mining:

Association Rule Mining is essential for unsupervised Market Basket Analysis. It helps discover product associations and co-purchase behaviours. Algorithms [5] like Apriori, Eclat and FP-Growth explore transactional data to find important item combinations. The metrics like support, confidence and lift can be used to validate the reliability and usefulness of the pattern found. This process enables businesses to improve shelf layouts, create promotional bundles and enhance cross-selling strategies, all without needing labelled datasets.

2.2 Algorithms and Enhancements:

Over the years, several algorithms and improvements have been created to address the limitations of basic Association Rule Mining methods. Apriori, one of the earliest and best-known algorithms, uses a level-wise approach [6] to generate candidate itemsets but it is costly in terms of computation because it requires multiple database scans and suffers from combinatorial explosion. To solve these problems, the FP-Growth [7] algorithm was introduced. It avoids candidate generation completely by using a compact prefix-tree structure known as the FP-tree. This approach greatly reduces both memory usage and runtime. Building on these ideas, newer methods have been suggested for more complex data situations.

2.3 Applications and Limitations:

Unsupervised Market Basket Analysis is widely used in the retail industry because it reveals patterns in customer buying behaviour without needing labelled outcomes. One of its main uses is in cross-selling and product bundling. Retailers identify items that are often bought together and use this information to group or promote products strategically. This method allows businesses to arrange products according to actual buying habits of customer to boosts sales. It is also useful for loss-leader [1] analysis where retailers use certain low margin items to attract customers in hope that they will also buy higher margin products.

The unsupervised Market Basket Analysis also has its limitations. As the number of items in a dataset grows, scalability becomes a major issue. For algorithms like Apriori [7] that need multiple scans of the database gives poor performance as the size of the database grows. It can also generate many redundant or trivial rules which leads to information overload. Another major limitation is its lack of personalization. The rules reflect general patterns instead of individual customer preferences. It makes this method less effective for targeted marketing. This method mainly describes past behaviour rather than predicting future actions. So it has limited use in dynamic and personalized retail marketing.

3. Supervised Learning in Market Basket Analysis

Supervised learning introduces a predictive aspect to Market Basket Analysis. Instead of identifying frequent itemsets, it builds models that can forecast customer behaviour using labelled data. Unlike unsupervised methods that look at general buying patterns, supervised learning finds the answer to specific questions such as "Will this customer buy product X?" or "What is the chance of purchasing item Y if items A and B are bought?" This method is useful for personalized marketing, inventory forecasting and dynamic recommendation systems.

3.1 Methodology and Algorithms

The supervised learning process in Market Basket Analysis starts by converting transactional data into a binary incidence matrix [3]. Each row represents a transaction and each column corresponds to an item. If an item appears in a transaction, the matrix shows a '1'; if not, it shows a '0'. A target label is also assigned, usually indicating whether a specific product was purchased. This type of dataset allows for various supervised machine learning algorithms to be used.

Decision Trees [8] and Random Forests [9] algorithms can analyse the presence of certain items and predict the chances of others being purchased. They provide clear models that are easy to understand, making them suitable for categorical transaction data. Logistic Regression [10] estimates the likelihood of a customer buying a specific item, which works well for binary classification tasks. Neural Networks [11] are good at modelling complex, nonlinear relationships among items and customer preferences.

Supervised learning provides many useful capabilities. For example, retailers can classify and predict product purchases, forecast quantities or customer spending [12]. They can also rank or recommend products based on individual users. This predictive ability of models helps businesses go beyond basic analytics. It allows for making strategies like personalized promotions, targeted ads and inventory management based on expected demand.

3.2 Advantage and Challenges:

Supervised learning has several key advantages over traditional unsupervised methods in Market Basket Analysis. It is effective at generating predictive and personalized insights. One of its main strengths is handling complex relationships among items. Whereas association rule mining usually focuses on simple co-occurrences, supervised models can understand relationships among multiple products. This leads to a better understanding of customer behaviour.

Supervised Market Basket Analysis can utilize information from external data sources like customer demographics or location. This improves the relevance of the analysis. It can also suggest the products based on individual preferences that can boosts customer engagement.

The supervised methods also come with challenges. A main requirement of this method is having labelled target outcomes [3]. For example, knowing in advance whether a specific item was purchased is necessary for predictive modelling to proceed. These models also may require extensive preprocessing like data cleaning and feature selection. They need more computational resources compared to unsupervised techniques. Another issue is the risk of overfitting while training on small or imbalanced datasets. In these cases, models may perform well on training data but fail on new, unseen transactions. Overall, while supervised models are more complex to implement, they provide a strong framework for predictive capabilities in market basket analysis.

4. Comparative Analysis:

Following Table – 1 shows the comparison of supervised and unsupervised methods for market basket analysis based on various aspects. This comparison shows the unique features and changing abilities of both methods in Market Basket Analysis. Supervised methods require more preprocessing but able to provide predictive insights that unsupervised models cannot. For large scale pattern discovery, unsupervised techniques are more suitable and computationally feasible.

Aspect	**Unsupervised MBA**	**Supervised MBA**
Objective	Discover frequent co-occurrences	Predict outcome (e.g., purchase)
Input	Transaction data	Transaction + target labels
Output	Association rules	Predictive model
Algorithms	Apriori, FP-Growth, Eclat	Logistic Regression, Trees, PRIM
Scalability	Moderate (depends on algorithm)	Higher with optimized algorithms
Interpretability	High (rules are explicit)	Varies (trees are interpretable)
Personalization	Poor	High

5. Hybrid, Fuzzy and Associative Classification Methods

As retail data becomes more complex, hybrid approaches have emerged to improve the effectiveness of Market Basket Analysis. These methods combine different techniques to overcome some limitations of traditional models. Traditional models depend on binary assumptions, lack context and struggle with scalability. Hybrid approaches give more flexible and intelligent mechanisms.

5.1 K-Apriori with Clustering

The K-Apriori [13] algorithm combines clustering and association rule mining. It uses K-Means clustering to group customers or transactions based on similarity before generating rules. By identifying similar subgroups, the algorithm allows rules to be mined within each cluster instead of across the entire dataset. This results in more relevant and specific association rules. This method is useful in customer segmentation, where different groups may show distinct purchasing behaviour. It supports targeted marketing strategies and personalized promotions.

5.2 Fuzzy Logic Integration

Traditional Market Basket Analysis techniques work based on the assumption that an item is either purchased or not. However real-world shopping behaviour of customers may have a different level of interest in a product. These uncertain, vague, or continuous data can be effectively handled by Fuzzy logic. Fuzzy logic [14] is useful in areas where item boundaries are not clearly defined. Fuzzy logic improves market basket models by addressing imprecision and uncertainty. The algorithms like Fuzzy Apriori [14] and Fuzzy FP-Tree [15] expand classical methods to manage continuous, vague, or linguistic data, such as "frequently," "occasionally," or "high quantity." These fuzzy models better reflect complex customer behaviour making them suitable for situations where strict thresholds do not apply. This method requires more computational power and complex data preprocessing that limits the scalability in very large datasets.

5.3 Association Classification

Another hybrid strategy is Classification Association Rule Mining. It combines the Association Rule Mining with the predictive capability of classification algorithms. The associative classification models like CPAR (Classification Based on Predictive Association Rules) [16] and CBA (Classification Based on Association Rules) [17] generate item association rules and also use them to predict class labels or outcomes. These algorithms can describe patterns and make predictions at the same time. This makes them ideal for applications that need both clarity and precision.

6. Challenges and Research Directions

One major challenge in Market Basket Analysis is the difficulty in managing large and sparse datasets. Retail databases can contain a large number of unique items but each transaction usually includes only a small selection. This results in a sparse matrix that presents challenges for both memory and computation. Another big challenge is filtering out redundant or irrelevant rules in unsupervised models like Apriori. These types of rules can generate thousands of patterns with a little actionable value. In supervised method, it is too difficult to include customer personal information like age, gender, location or preferences or how recently a purchase was made into traditional models. This requires lot of preprocessing and changes to the models.

Several research directions aim to handle these challenges and expand the capabilities of Market Basket Analysis. Deep Learning [18] techniques provide new ways to represent products as vectors in latent space. These vectors capture complex relationships between items that traditional models might miss. Real-time Market Basket Analysis using data stream mining is another emerging trend that is useful for e-commerce platforms where transaction data is generated continuously and decisions must be made quickly. To enhance the understanding of item associations, ontology-based Association Rule Mining [19] uses domain knowledge and hierarchical relationships between products.

7. Conclusion

Market Basket Analysis has changed from simple association mining to more advanced predictive modelling. Unsupervised methods like association rule mining are still useful for exploratory analysis, while supervised learning allows for predictive retail analytics. Fuzzy and hybrid methods improve the ability to handle complex, uncertain or structured data. The decision on which methods to use should depend on the type of data, business goals and computing limits. A combination or multi-stage approach can be used to achieve the best results by finding patterns with unsupervised tools and validating them with supervised models.

References

References

[1] H. Aguinis, L. E. Forcum and H. Joo, "Using market basket analysis in management research," *Journal of Management,* vol. 39, p. 1799–1824, 2013.

[2] A. A. Raorane, R. V. Kulkarni and B. D. Jitkar, "Association rule–extracting knowledge using market basket analysis," *Research Journal of Recent Sciences ISSN,* vol. 2277, p. 2502, 2012.

[3] G. Kronberger and M. Affenzeller, "Market basket analysis of retail data: supervised learning approach," in *International Conference on Computer Aided Systems Theory*, 2011.

[4] C. S. Deora, S. Arora and Z. Makani, "Comparison of interestingness measures: support-confidence framework versus lift-irule framework," *IJERA International Journal of Engineering Research and Applications,* vol. 3, 2013.

ISSN 2347–3657

Volume 6, Issue 1, 2018

[5] S. Solanki and N. Soni, "A survey on frequent pattern mining methods Apriori, Eclat, FP growth," *International Journal of Computer Techniques,* vol. 10, p. 86–100, 2013.

[6] J. Dongre, G. L. Prajapati and S. V. Tokekar, "The role of Apriori algorithm for finding the association rules in Data mining," in *2014 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT)*, 2014.

[7] A. K. Singh, A. Kumar and A. K. Maurya, "An empirical analysis and comparison of apriori and FP-growth algorithm for frequent pattern mining," in *2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies*, 2014.

[8] J. Fürnkranz, "Decision tree," in *Encyclopedia of machine learning*, Springer, 2011, p. 263–267.

[9] A. Cutler, D. R. Cutler and J. R. Stevens, "Random forests," *Ensemble machine learning: Methods and applications,* p. 157–175, 2012.

[10] A. DeMaris, S. H. Selman, A. DeMaris and S. H. Selman, "Logistic regression," *Converting data into evidence: A statistics primer for the medical practitioner,* p. 115–136, 2013.

[11] A. D. Dongare, R. R. Kharde, A. D. Kachare and others, "Introduction to artificial neural network," *International Journal of Engineering and Innovative Technology (IJEIT),* vol. 2, p. 189–194, 2012.

[12] A. Raorane and R. V. Kulkarni, "Data mining techniques: A source for consumer behavior analysis," *arXiv preprint arXiv:1109.1202,* 2011.

[13] L. C. M. C. Annie and A. D. Kumar, "Market basket analysis for a supermarket based on frequent itemset mining," *International Journal of Computer Science Issues (IJCSI),* vol. 9, p. 257, 2012.

[14] C. M. Kuok, A. Fu and M. H. Wong, "Mining fuzzy association rules in databases," *ACM Sigmod Record,* vol. 27, p. 41–46, 1998.

[15] C.-W. Lin, T.-P. Hong and W.-H. Lu, "Linguistic data mining with fuzzy FP-trees," *Expert Systems with Applications,* vol. 37, p. 4560–4567, 2010.

[16] X. Yin and J. Han, "CPAR: Classification based on predictive association rules," in *Proceedings of the 2003 SIAM international conference on data mining*, 2003.

[17] N. Abdelhamid and F. Thabtah, "Associative classification approaches: review and comparison," *Journal of Information & Knowledge Management,* vol. 13, p. 1450027, 2014.

[18] X.-W. Chen and X. Lin, "Big data deep learning: challenges and perspectives," *IEEE access,* vol. 2, p. 514–525, 2014.

[19] I. N. Ferraz and A. C. B. Garcia, "Ontology in association rules," *SpringerPlus,* vol. 2, p. 1–12, 2013.