Approach to Kirana Store Product Arrangement Using Machine Learning

Ajita Patel1* and Krishna Kumar Tiwari1

1REVA Academy for Corporate Excellence, REVA University, Rukmini Knowledge Park, Kattigenahalli, SH 104, Srinivasa Nagar, Bengaluru, Karnataka 560064, India.

Authors‘ contributions

This work was carried out in collaboration between both authors. Both authors read and approved the final manuscript.

Article Information

DOI: 10.9734/AJRCOS/2021/v11i430271

Article Information

DOI: 10.9734/AJRCOS/2021/v11i430271

Editor(s):
(1) Prof. G. Sudheer, GVP College of Engineering for Women, India.

Reviewer(s):
(1) Jyoti Dabass, The Northcap University, India.
(2) Satyasis Mishra, Adama Science and Technology University, Ethiopia.

Complete Peer review History: https://www.sdiarticle4.com/review-history/73667

Received 29 June 2021
Accepted 09 September 2021
Published 15 September 2021

Original Research Article

ABSTRACT

Market Basket Analysis (MBA) is a method for determining the association between entities, and it has often been used to study the association between products in a shopping basket. Trained Computer vision models are able to recognize objects in photos so accurately that it can even outperform humans in some instances. This study shows that combining objective detection techniques with market basket analysis can assist Stores/Kirana in organizing the products effectively. With the use of MBA and Object detection, we formulated recommendations for store arrangements along with putting a recommendation engine on top to help shoppers. After deploying this to local Kirana stores, the Kirana store was able to see an increase of 7% in the sale. The recommendation engine performed better than just the domain knowledge of the kirana store.

Keywords: Computer vision; object detection; market basket analysis; association rules.

1. INTRODUCTION

Millions of small shops known as Kirana stores make up the largest part of retail trade, which contributes 11 percent to GDP (secondary to agriculture with 18 percent). Food and grocery make up more than 65 percent of the retail sector in India [1]. However, the majority of stores...
continue to use the same old infrastructure, which includes a shabby appearance and unorganized items. On the other hand, the Consumer attitude and mindset are shifting towards a more clean and organized infrastructure. It is critical for products to be visible for customers to purchase them.

Today's business is greatly impacted by globalization. The growing number of supermarkets, which means a larger number of clients, has sparked interest in the invention and analysis of search algorithms [2]. Since 1980, it has been increasingly popular to search for hidden data in a supermarket's database. Businesses must find a way to thrive in a competitive environment and demonstrate their relevance. The best strategy to survive in this challenging environment is by knowing our customers and their preferences [2]. In business, data mining is critical for making smart business decisions and understanding consumer behavior. When a Market Basket Analysis is used, it is capable of identifying the relationships between buyers’ purchasing patterns and Association Rule Mining, which is capable of identifying the relationships between different products a buyer purchased [3].

Market Basket Analysis is a well-known and widely used tool by major retailers to uncover relationships between items. It helps in finding consumer’s shopping trends by collecting associations or co-occurrences from transactional data from a shop [4]. It operates by looking for a combination of items that sometimes appear together in exchanges. In other words, it enables retailers to understand associations between items that customers buy. In competitive markets, businesses must turn this data into valuable information and insight for decision-making. The goal of association rule mining is to find associations between things customers have bought in transaction databases. For transactional databases, association rules are expressed in the form X ->Y, where X and Y denote sets of elements. A rule like this implies that customers who buy items in X tend to also buy items in Y [5]. If marketers know that consumers who buy one product are likely to buy another, they can sell the two products together or make target customers for the second product [6].

Placement of products on the sales shelves does not currently reflect the needs of consumers, as it is mostly determined by management's perceptions and not captured from the consumer's perspective [7]. To create a better store layout, it becomes important for the retailer to know what items are available in the store. Managing a good shelf space will not only decrease inventory level but also give customer satisfaction as well. The automatic detection of objects on shelves can easily handle several difficulties in the market, such as recognizing products or brands, keeping track of the products on the shelves, or filling in the gaps [8]. Currently, most supermarkets employ staff to check the tally of the shelf, which is inaccurate, time-consuming, and incurs high labour costs. In this case, computer vision technology must be used to detect the object of supermarket products against a background of shelves [9]. For determining the association rules, the project uses the automatically detected items in the shelves. It makes sense to organize items according to the most frequently purchased and connected items to make the sale and service of goods more efficient, helpful, and efficient.

2. LITERATURE REVIEW

The competitive nature of the market always makes it hard for companies to keep a good position because winning the market relies on the ability to make good decisions and understand customer behavior. Studying the customer purchasing patterns is critical to a business [3]. One of the best and most well-known examples of data mining association rules is Market Basket. Many algorithms are developed by researchers to help users achieve their goals. Market basket analysis, also known as affinity analysis, is a modelling tool that aids in identifying commodities that are likely to be purchased together by observing a typical item set. It assumes that the basket contains a large number of products. Each customer buys a subset of products that suit their needs, and advertisers learn which products have been purchased. Advertisers use this data to place items at different locations [3]. Agrawal, Imieliński, and Swami [10] proposed the Apriori algorithm. They analyzed previously collected transaction information from consumers to uncover the patterns of association among the items purchased. In a short period, it became the standard method for several real-world marketing applications.

Researchers Umami & Surjandari [4] studied the importance of Place, Promotion, and Price (4Ps) in their research. To gain a competitive edge,
they determined how to use the 4Ps. Ünvan and Yüksel discussed the conviction value. According to their research, if the conviction value is 1, the products are usually independent of one another. Therefore, it is not considered as a rule [11].

2.1 Methods of Market Basket Analysis

One of the most widely used methods of data analysis in marketing is Market Basket Analysis (MBA). The analysis is conducted to find out what products customers purchase or use most often. MBA is a good example of how the Association Rule is put into practice.

2.1.1 Association rule mining

As the foundation for market basket analysis, association rule mining (ARM) identifies the relationships or associations between large sets of data items [4]. ARM rules can be broken down into Frequent Itemset, Utility Itemset, and Rare Itemset mining [12]. Fig 1 displays the different types of association rule mining.

An association rule expresses the idea of “what goes with what”. Transactions refer to the purchase of products made by consumers at markets. Three parameters can determine the size of an associative rule i.e. support, confidence and lift [13].

- **Support**

Information about how many times items have appeared in overall transactions as displayed in equation 1 is called Support.

\[
Support = \frac{\text{Total number of Transactions that contains } A}{\text{Total Transactions}}
\]  

(1)

- **Confidence**

A measure of a customer's likelihood of buying product A after buying product B. An association rule would be expressed as (items from set A) to (items from set B), where A would be the precedents and B would be the consequences as displayed in equation 2. Based on pre-existing antecedents, confidence tells us how likely Consequence is to appear on a cart.

\[
Confidence = \frac{\text{Total Number of Transactions that Contains } A \text{ and } B}{\text{Total Number of Transactions that Contains } A}
\]  

(2)

- **Lift**

The lift of A=>B is measured by dividing the confidence of A=>B by the support of B as displayed in equation 3. Lift is symmetric. That is, the lift of A=>B has the same value as the lift of B=>A. If the lift of A=>B is less than 1, then the occurrence of A is negatively associated with the occurrence of B. If the resulting value is greater than 1, then A and B are positively correlated [5].

\[
Lift = \frac{\text{Confidence}(A \Rightarrow B)}{\text{Support}(B)}
\]  

(3)

2.1.2 Apriori algorithm for finding frequent itemset

The Apriori algorithm evaluates a data set to decide which items are most likely to occur together. Apriori algorithm [3] is the most commonly used algorithm for finding the association rules of baskets. Apriori algorithm [14] follows two main processes:

Merge: A merge is a process in which items are combined so that no further combinations are possible.

Prune: Using the minimum support, the user prunes the results of the combined items.

2.1.3 Frequent Pattern Growth Algorithm

FP Growth uses an FP Tree graphic data structure to compress transactional records in order to measure frequent article collection. FP-Tree can be imagined in such a way that data records are converted into a graph format. In contrast to the Apriori algorithm, which used generation and checking to generate the regular itemset, FP-Growth first generates the compact FP tree, from which a regular itemset is derived.

To search for item sets, a prefix tree is built or a FP tree is built. The infrequent elements are pruned from the tree. Similar to the Apriori algorithm, pruning is calculated based on the minimum support value. After that, a set of rules are generated based on the minimum confidence value [12]. As described by Unvan, Fig. 2 shows some of the differences between Apriori and FP Growth.

2.2 Object Detection Algorithm

The rise of deep learning led to the development of convolutional neural networks (CNNs) based solutions to object detection issues. In object recognition, objects are identified and classified in an image, and their existence is indicated by rectangular bounding boxes. The sliding window
method was the most popular method for locating and classifying pipelines in the past.

For addressing the slowness of the R-CNN model, the Fast R-CNN model was developed in 2015. There are many similarities between this method and R-CNN, but it is faster than R-CNN [8]. However, in 2016, a research group, in partnership with Facebook AI, proposed training the entire image with a ConvNet rather than choosing parts and running them separately. This enhanced both the accuracy and the time it took to detect the object. It’s also known as “You Only Look Once” (YOLO)” [15]. Comparison of a number of object recognition algorithms proved that YOLO performed most effectively for recognizing items on store shelves [8].

3. METHODOLOGY

In this research, CRISP-DM, known as Cross-Industry Data Mining, is used. CRISP-DM is a structured method to plan and execute data mining projects. Fig.3 illustrates the CRISP-DM framework.
4. BUSINESS UNDERSTANDING

The Kirana shop is located in a residential area of Bangalore, India surrounded with many apartments. It has a carpet area of approximately 950 square feet. It sells vegetables, fruits, and grocery items including Dal, Rice, Milk, Cooking Oil etc. It has an average daily footfall of 100-150 customers. The store’s billing system is computerized, and it accepts credit/debit card payments as well as UPI payments. Online retailers such as Bigbasket, Grofers, Reliance Fresh, and nearby supermarkets have been putting a lot of pressure on it in recent years. Consequently, the Kirana shop needs to rethink the way they do business in order to not only survive the competition but also tap into consumers’ demands for new buying methods.

5. DATA UNDERSTANDING

Two datasets are included in the project e.g.: Image dataset and Customer Transactional Dataset. The image data was collected from various local markets using a mobile phone and some of the images were taken from publicly available sources [16]. This data consists of common vegetables and fruits, such as tomatoes, potatoes, onions, lemons, bananas, etc. This image dataset contains 824 images from 12 different classes.

Customer Transaction data was collected from the Kirana store, Point of Sale (PoS) was collected from customers and data also taken from publicly available sources [17]. The PoS were scanned first and the data was extracted using an OCR reader.

6. DATA PREPARATION

The images were labelled using LabelImg tool before they were further processed. As part of YOLO object recognition, each image should have a defined boundary box and its class name. Labels were created in the Pascal VOC XML format as shown in Fig 4. The size of the dataset is further increased in this step, by rotating and adding noise to the images, shifting them, mirroring them, and adding filters using image augmentation.

The transactional data collected from the customer was cleaned and transformed into a suitable .csv format for further analysis. Fig. 5 shows a sample data.

7. DATA MODELLING

As presented in Fig. 6, the first aim of the model is to identify the objects. To accomplish this goal, the object identification model needs to get high accuracy with minimal loss. YOLOv4 Tiny custom model needs a powerful GPU for training. Google Colab already provides a free high-performance GPU, which is the Tesla K8 with up to 12 GB. A transfer-learning model is used to speed up training or increase the depth of learning by transferring learning from one role to another algorithm’s performance. The training of the object detection issue was done using previously trained YOLO convolutional weights.

The mAP ((Mean Average Precision) is the most common term in deep learning to evaluate models. It calculates based on how well your prediction is. Calculation of the mAP is based on Precision, Recall, IoU (Intersection over Union). IoU presents the accuracy of the predicted bounding boxes. Table 1 shows the model score.

Utilizing the ‘Apriori’ function of Mlxtend to determine together which items are most commonly purchased. The ‘Apriori’ role requires minimal support in order to receive frequent item sets. When low support levels are used, frequent itemsets can have very few results, and when high support levels are used, it takes a lot of memory to process the data. The support value is kept at 0.02 for this analysis.

The next step is to use Mlxtent’s ‘association rules’ function to build the association rules. Choose the most relevant parameter (either lift or confidence) to determine the minimum threshold (called the min threshold).

Different association rules are generated by varying the length of the items. First the rules are generated for two antecedents with one consequent (2 X 1) and then for three antecedents with one consequent (3 X 1).

Fig. 7 shows the most correlated items here are Banana, Cucumber =>Orange.

Fig. 8 shows the most correlated items are Root Vegetables, Brinjals, Beans => Gourd.

8. MODEL EVALUATION

Precision, Recall, F1-score, mAP (Mean Average Precision), IoU (Intersection over Union) of the test results are calculated for evaluation. YoloV4
Tiny gave a prediction of 92.5% on the test data set with an avg IoU of 70.6%.

For the association rules, the support vs confidence and lift vs confidence metrics are evaluated. The rules that lie on the right hand border of the plot where either support, confidence or both are maximized are picked as displayed in Fig. 9.

A positive relation is seen between the item sets A and B as displayed in Fig. 10.
Table 1. YoloV4 Tiny model score

| mAP     | Precision | Recall | F1-Score | Avg IoU  |
|---------|-----------|--------|----------|----------|
| 2.52%   | .90       | .90    | .90      | 70.63%   |

Fig. 7. For two antecedents with one consequent
Fig. 8. For three antecedents with one consequent
Fig. 9. Support Vs Confidence

Fig. 10. Lift Vs Confidence
Table 2. Top association Rules

| Support, Confidence | Associated Rule                                                                 |
|---------------------|---------------------------------------------------------------------------------|
| support: 0.1,       | ('Moong Dal')=>('Toor Dal', 'Cashews')                                         |
| confidence: 0.5     |                                                                                 |
| support: 0.01,      | ('Ghee', 'Moong Dal', 'Rice Flour', 'Whole Spices')=>('Sugar')                 |
| confidence: 0.5     | ('Brinjals', 'Cucumber', 'Gourd', 'Potato', 'Root Vegetables')=>('Onion')     |
|                     | ('Moong Dal', 'Cucumber', 'Gourd')=>('Banana')                                  |
|                     | ('Toor Dal', 'Brinjal', 'Cucumber')=>('Bhindi')                                 |
|                     | {'Besan', 'Moong Dal'}=>{'Toor Dal'}                                            |
|                     | {'Bread', 'Cucumber', 'Toor Dal'}=>{'Brinjals'}                                 |
| support: 0.01,      | ({'Cucumber', 'Brinjals', 'Bhindi'})=>{'Toor Dal'})                             |
| confidence: 0.7     |                                                                                 |
9. DEPLOYMENT

The object detection YoloV4 Tiny model is deployed using Flask and the UI for the recommendation was engine is created using PHP. The vendor takes a picture of shop and uploads to the object detection model. The model detects the items in the picture with a trained object detection model. The vendor picks up the items and look forward to the associated items from the items available in the picture. Fig. 11 shows the object detection and the Recommendation frontend.

10. RESULTS AND INSIGHTS

Different association rules are generated by varying the confidence and support levels. When the confidence and support are high, only a few results are generated whereas when lowered the values more rules were generated. Only the top few selected rules are presented in Table 2.

11. CONCLUSION

During this study, the goal was to identify the items available in the Kirana store and the customer-buying pattern. For example, there is a high chance that when a customer buys Toor Dal, the customer will also going to buy Cucumber, Brinjals and Bhindi. The vendor can make sure these items are available in the store and then this data can be utilized for future planning for the store layout design and for inventory planning. This research only has a few months of data. Additional data must be collected for future research. As a result, the results would be more accurate.

DISCLAIMER

The products used for this research are commonly and predominantly use products in our area of research and country. There is absolutely no conflict of interest between the authors and producers of the products because we do not intend to use these products as an avenue for any litigation but for the advancement of knowledge. Also, the research was not funded by the producing company rather it was funded by personal efforts of the authors.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Retailers TG. Transforming Kirana Stores To Drive Economic; 2021. Available:https://www.accenture.com/in-en/insights/local/kirana-transformation
2. Pukach P, Shakhovska K. The mathematical method development of decisions supporting concerning products placement based on analysis of market basket content. 2017 14th Int Conf Exp Des Appl CAD Syst Microelectron CADSM 2017 - Proc. 2017;347–50.
3. Dubey SK, Mittal S, Chattani S, Shukla VK. Comparative Analysis of Market Basket Analysis through Data Mining Techniques. Proc 2nd IEEE Int Conf Comput Intell Knowl Econ ICCIKE. 2021; 239–43.
4. Utami NT, Surjandari I. Identifying consumer buying behavior differences through market basket analysis in multiple outlet types. ACM Int Conf Proceeding Ser. 2017;Part F1319:82–6.
5. Ahn K Ii. Effective product assignment based on association rule mining in retail. Expert Syst Appl [Internet]. 2012;39(16):12551–6. Available:http://dx.doi.org/10.1016/j.eswa.2012.04.086
6. Nichol MB, Knight TK, Dow T, Wygant G, Borok G, Hauch O, et al. Quality of anticoagulation monitoring in nonvalvular atrial fibrillation patients: Comparison of anticoagulation clinic versus usual care. Ann Pharmacother. 2008;42(1):62–70.
7. Yang MH, Chen WC. Study on shelf space allocation and management. Int J Prod Econ. 1999; 60(510):309–17.
8. Melek CG, Sonmez EB, Albayrak S. Object Detection in Shelf Images with YOLO. EUROCON 2019 - 18th Int Conf Smart Technol. 2019;1–5.
9. Geng Z, Wang Z, Weng T, Huang Y, Zhu Y. Shelf Product Detection Based on Deep Neural Network. Proc - 2019 12th Int Congr Image Signal Process Biomed Eng Informatics, CISP-BMEI; 2019.
10. Agrawal R, Imielinski T, Swami A. Mining Association Rules Between Sets of Items in Large Databases. ACM SIGMOD Rec. 1993;22(2):207–16.
11. Ünvan YA. Market basket analysis with association rules. Commun Stat - Theory Methods [Internet]. 2021;50(7):1615–28. Available:https://doi.org/10.1080/03610926.2020.1716255
12. Isa N, Kamaruzzaman NA, Ramlan MA, Mohamed N, Puteh M. Market Basket Analysis of Customer Buying Patterns at Corm Café. Int J Eng Technol. 2018;7(October):119–23.

13. Gurudath S, Analysis MB. Market Basket Analysis & Recommendation System Using Association Rules Market Basket Analysis & Recommendation System Using Association Rules Shruthi Gurudath Submitted in partial fulfillment for the degree of Master of Science in Big data management and. 2020;(August).

14. Widodo ID, Ulfah H, Anggraeni KN. Redesign Super Market Layout Analysis Based on Hidden Customer Purchase Behaviour. 2021 IEEE 8th Int Conf Ind Eng Appl ICIEA. 2021;261–4.

15. Vellala AK. Genre-based Video Clustering using Deep Learning By Extraction feature using Object Detection and Action Recognition. Div Stat Mach Learn [Internet]; 2021. Available: http://liu.diva-portal.org/smash/record.jsf?pid=diva2%3A1570452&dswid=-8178

16. Shaikh A. common vegetable dataset for object detection; 2020. Available: https://www.kaggle.com/slrq123/common-vegetable-dataset-for-object-detection

17. Mohammad Imran Shaikh. Big Basket Dataset [Internet]; 2020. Available: https://www.kaggle.com/general/181642. Available: https://www.kaggle.com/general/181642

© 2021 Patel and Tiwari; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle4.com/review-history/73667