International Journal for Innovative Engineering and Management Research

A Peer Revieved Open Access International Journal

www.ijiemr.org

IJIEMR Transactions, online available on 27th Nov 2022. Link

:http://www.ijiemr.org/downloads.php?vol=Volume-11&issue=Issue 11

DOI: 10.48047/IJIEMR/V11/ISSUE 11/24

Title **CUSTOMISED MARKET BASKET ANALYSIS USING MACHINE LEARNING**

Volume 11, ISSUE 11, Pages: 149-153

Paper Authors

B. Shiva Sahu, Dr. G. Venkata Rami Reddy

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

CUSTOMISED MARKET BASKET ANALYSIS USING MACHINE LEARNING

B. Shiva Sahu
*Student, M. Tech (computer Science), School of Information Technology, JNTU Hyderabad, Hyderabad, India
Dr. G. Venkata Rami Reddy
*Professor of IT, School of Information Technology, JNTU Hyderabad, Hyderabad, India

ABSTRACT:
Offering customized services to clients is a hot topic these days for supermarket chains. One of these services is market basket prediction, which involves providing the customer with a shopping list for their future buy based on her present needs. The many elements influencing the customer's decision process: co-occurrence, sequentially, periodicity, and recurrency of the purchased items are not captured simultaneously by current techniques. To do this, i created a pattern called Temporal Annotated Recurring Sequence (TARS), which can capture all of these factors simultaneously and adaptively. Denifed a method for extracting TARS and create TBP (TARS Based Predictor), a predictor for the next basket that, in addition to TARS, can understand the next basket. And also i will send a mail to customer which contains the data of their past purchases where they get an idea how items they bought many times and in which season they go for selected products. So, this will give an overview for their customers regarding their purchases.

Keywords- *Next basket prediction, temporal recurring sequences, user-centric model, market basket analysis, data mining, interpretable model.*

1. INTRODUCTION

Effective marketing strategies and engagement tactics face a significant problem in identifying purchasing patterns and how they change over time. Basket prediction, or the automated foretelling of a customer's next basket purchase, is one of the most promising services that retail markets may provide to their clients in this context. A good basket recommender can serve as a shopping list reminder by recommending products that the customer might require.

This must be realised successfully, and that demands a thorough understanding of a person's shopping habits. Individuals' shopping habits change over time for a variety of reasons, including dietary changes or shifts in personal tastes, as well as environmental factors like seasonality of items or store rules.

As a result, a successful solution to basket prediction must be flexible enough to take into account how a client behaves over time, how her buying habits are repeated, and how they alter on a regular basis. I suggest the Temporal Annotated Recurring Sequences (TARS), adaptable patterns that represent a person's shopping behaviour using four key factors. TARS first takes into account the co-occurrence: when a consumer routinely buys a group of things together. The sequentiality of purchases, or the fact that a consumer routinely buys a group of things one after the other, is the second aspect of TARS that is modelled. Third, TARS takes periodicity into account. For either environmental or private reasons, a customer can only systematically make a series of purchases at certain times of the year. Fourth, TARS takes into account how frequently a sequential purchase occurs during a customer's period of the year, or its recurrency during each period. Co-occurrence, sequentiality, periodicity, and recurrence are

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

the four aspects that must be modelled in order to identify a person's buying behaviour and track its development in time. On the one hand, what a consumer will buy depends on what they already bought. Future needs depend on demands now satisfied. On the other hand, a customer's needs depend on her unique behaviours, or the repeated purchases she makes time and time again. Shopping is not a static activity. Both endogenous and individual factors have an impact on habits. Periodicity is therefore a key feature of an adaptive model for basket prediction. Here using TARS to build a parameter-free TARS Based Predictor (TBP) that solves the basket prediction issue and outputs a recommended basket as a list of goods to keep in mind while making the next purchase. So I designed a code to send a mail to individual customers where I get that mail during billing time so the customers get an idea how their purchases changes according to the time and seasons. So this project help Vendor as well as Customer to get a clear cut idea of how they are selling and purchasing the products respectively.

2. LITERATURE SURVEY

One of the primary marketing topics that analysts and brand executives from firms have recently been concentrating on is repurchasing. In contrast to prior user-retailer partnership research, which focuses on the behaviours of each seller or customer, this essay analyses the idea of relational phenomena using an integrative methodology that takes into account the traits and attitudes of both parties.The purpose of using the relationship hypothesis to demonstrate customer purchasing behaviour is to raise awareness of the interactions between businesses and customers. The foundation for the issues presented in this paper was developed by the Department of Business Studies at Uppsala University using a thesis plan and an application for a larger research project. Four senior employees and three PhD candidates worked on the research project.

2.1. Temporal Annotated Recurring Sequences:

Two aspects are modelled by Temporal Annotated Recurring Sequences: I the customer's recurring and sequential purchases, that is, the common practise of buying a group of things together and following another group of items; (ii) the occurrence of the sequential purchase, or when and how frequently such a pattern appears in the customer's purchasing history. With the use of a real-world example,i explain their components and make clear their significance in order to demonstrate how TARS captures these two elements simultaneously.

2.2. TARS extraction procedure:

Here applying an extension of the well-known FP-Growth method to extract the TARS from a customer's purchasing history. i choose FP-Growth despite the fact that a number of algorithms can be utilized to do the same objective. First off, FP-Growth produces straightforward outcomes .In order to get a more accurate representative value, utilized the median to aggregate the number of occurrences in each period and as the aggregation function in Algorithm 2. In fact, the median value is a useful representative value for skewed distributions as well since it is less vulnerable to potential outliers than the mean.

Due to two main factors, i only take into account sequences with two item sets and not sequences with more item sets. The first one is a result of the definition of TARS: since i wish to utilize them for prediction, this modelling simply requires a head and a tail to be used for determining the rank of the items.

2.2.1. Data-Driven Parameters Estimation:

Here i perform two pre-processing procedures on the base to make the parameters adaptive to the sequences as well as the particular customer. Data-driven estimation of the sets of parameters outlined in the algorithm constitutes the initial stage of pre-processing.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Assume S to be the set of base sequences that fall within the median of the inter-times in DS. i calculate parameter from a base sequence S. i put base sequences that have similar intertimes together to create a set of clusters.Then,i estimate and group the base sequences with similar median number of occurrences each period, resulting in a set of clusters, and calculate the periods TCS compatible solely with the temporal constraint.

3.TARS BASED PREDICTOR

The TARS Based Predictor, an approach for market basket prediction that is noticeably personalised and user-centric, is built on top of the set of TARS extracted from a customer's purchase history. Predictions for a customer are made using only the model built on her purchase history, i.e., her TARS.

TBP takes advantage of TARS to simultaneously encode sophisticated item interactions such as co-occurrence (which goods are purchased together), sequential relationship (which things are purchased after), periodicity (which items are purchased when), and typical repurchase intervals (after when re-purchases happen).These elements give TBP the ability to examine a customer's recent purchase history and identify the active patterns, or the buying habits that the customer is now engaging in. In turn, by being aware of active patterns, TBP is able to supply the goods the consumer will require while making their subsequent purchase. It is important to note that TBP is parameter-free: all the TARS model Gc's parameters are automatically estimated for each customer based on their particular data BC, avoiding the typical situation where the same parameter setting is applied to all customers uniformly.

4. METHODOLOGY

The distinction between OOA and OOD becomes more ambiguous when object orientation is employed both in analysis and design. This is especially valid for methodologies that incorporate both analysis and design. The similarities between the fundamental building blocks—classes and objects—used in OOA and OOD are one cause of this blurring. There is some general agreement regarding the realms of the two activities, even though there is disagreement over which aspects of the object-oriented development process belong to analysis and which parts to design.

The key distinction between OOA and OOD is that the former models the problem domain, resulting in a comprehension and specification of the problem, while the latter models the problem's resolution. In other words, analysis addresses the problem domain, while design focuses on the problem-solving area. However, with OOAD, the solution domain representation is included. That is, a large portion of the representation developed by OOA is typically present in the solution domain representation, which was created by OOD. Varied people have different perspectives on the dividing line because it is a question of perception. One of the advantages of the object-oriented method is that there is no obvious distinction between analysis and design; rather, the shift is "seamless." This is also the basic justification for OOAD approaches, which combine analysis and design. Due to the diverse modelling domains, the fundamental distinction between OOA and OOD is found in the kinds of objects that result from the analysis and design stages.

5. CODE

Part of code:

- import numpy as np
  # For Mathematical calculation we are importing Numpy
- import pandas as pd
  # For Loading the dataset we are importing pandas
- from mlxtend.preprocessing import TransactionEncoder
  # Using and TransactionEncoder object,
  we can transform this dataset into array format suitable for typical machine learning APIs.
- from mlxtend.frequent_patterns import fpgrowth
  # Applying FP Growth algorithm for taking frequent itemset.
- rom mlxtend.frequent_patterns import association_rules
  # applying association rule for data mining.
- import seaborn as sns
  # for Visualization we are importing seaborn, matplotlib, wordcloud, squarify.
- import matplotlib.pyplot as plt
- from wordcloud import WordCloud
- import squarify
- import networkx as nx

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

-sending email to customer:

```
# import smtplib
# s = smtplib.SMPT('smtp.gmail.com', 587)
# s.starttls()
# s.login("senderemail@gmail.com, "xxxxxx")
# message = "Recommended Basket for you\n" + finalBasket
# s.sendmail("senderemail@gmail.com", customer email, message)
# s.quit()
# print("mail sent to customer")
In[].
```

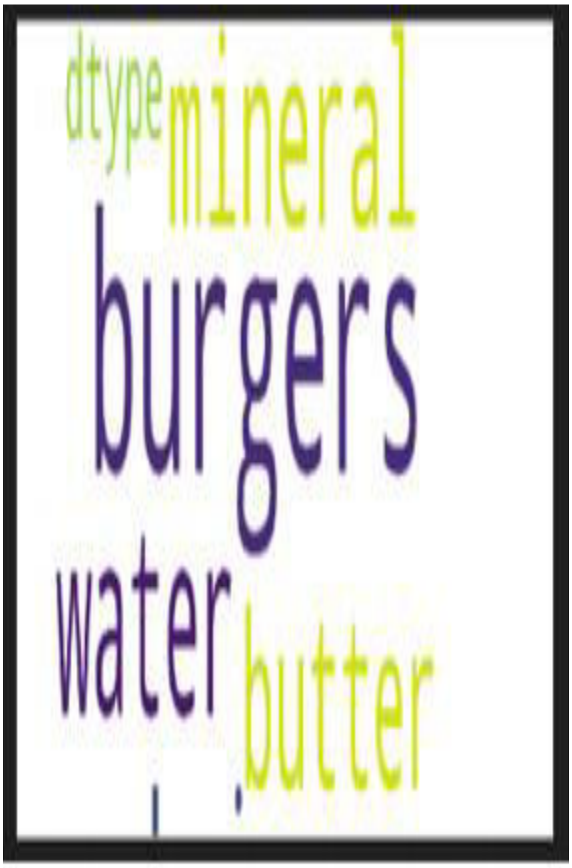

Fig.2: Live word clouds of popular items after the poll

6. OUTPUT SCREENS

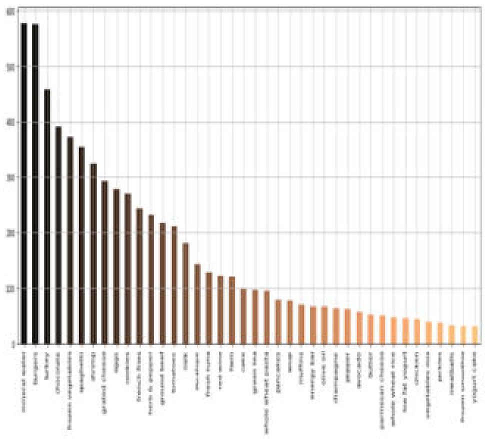

Fig.1: Frequency of most popular items

7. Conclusion

A method to market basket analysis emphasizing evidence, interpretability, and user-centricity is provided in this work. For the purpose of forecasting the next basket using Temporal Annotated Recurrent Sequences, i have described and used a TARS-based predictor. TBP uses the customer's action specificity to modify how TARS is extracted and so produces more customized models because it is parameter-free. Here an email also sent to individual customer regarding their purchase to know their every purchase and it will be easy for them when they visit again. I also conducted research using actual data sets to show that TBP uses cutting-edge techniques and also discovered that the TARS extraction provides helpful trends that can be used to learn more about customer purchasing patterns as well as product characteristics like seasonality and shopping seasons. According to the results, a customer's purchase behaviour takes at least 36 weeks to accurately predict his potential purses. In this scenario, TBP will forecast the subsequent 20 baskets very precisely. So this helps both vendor and customer to know their transactions. For a vendor they will come to know how their product purchases are made and with that analysis they make modifications to their business accordingly and for a customer, they gets an idea how their purchases are and what type of product they going to purchases based on time and seasons.

References

[1] S. Post, and al., “Factorizing personalized Markov chains for next-basket recommendation,” In Sec. In Admin. Nineteenth Ins. Conf.: Conf. World Wide Network, 2010, pp. 97-71.

[2] S. Set, and al., “BPR: Bayesian personalized ranking from implicit feedback,” In Sec. In Admin. Conf. 25th. Artif Uncertainty. Lot., 2009, pp. 452-461.

[3] R. Wang and others., “Learning hierarchical representation model for NextBasket recommendation,” In Sec. In Admin. Thirty-eight Ins. SIGIR ACM Conf. Conf. F.-Res. Build. Production. Inf. Data. Construction, 2015, pp. 403-412.

[4] F. Yu and others., “A dynamic recurrent model for next basket recommendation,” In Sec. In Admin. Int. 39th. SIGIR ACM Conf. Conf. F.-Res. Build. Production. Inf. Data. Print, 2016, pp. 729-732.

[5] R. Wang, among others., “Modeling retail transaction data for personalized shopping recommendation,” in Sec.-Proc. ACM Int. 23rd. Conf.: Conf. Inf. Data. Knowl.-Knowl. Technology., 2014, 1979-082. Finance

[6] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, et al., “A survey of methods for explaining black box models,” ACM Comput. Surv., vol. 51, no. 5, 2018, Art. no. 93.

[7] A. Pentland, et al., “Personal data: The emergence of a new asset class,” in An Initiative of the World Economic Forum, 2011.

[8] W. A. Kosters, et al., “Complexity analysis of depth first and FPgrowth implementations of APRIORI,” in Proc. Int. Workshop Mach. Learn. Data Mining Pattern Recognit., 2003, pp. 284–292.

[9] C. Kalapesi, “Unlocking the value of personal data: From collection to usage,” in World Economic Forum Technical Report, 2013.

[10] J. Dem——————————sar, “Statistical comparisons of classifiers over multiple data sets,” J. Mach. Learn. Res., vol. 7, no. Jan, pp. 1–30, 2006.

[11] A. Monreale, et al., “Privacy-by-design in big data analytics and social mining,” EPJ Data Sci., vol. 3, 2014, Art. no. 10.