



Fuzzy association rule mining approach to identify e-commerce product association considering sales amount

Onur Dogan^{1,2} · Furkan Can Kem¹ · Basar Oztaysi³

Received: 26 June 2021 / Accepted: 3 December 2021 / Published online: 4 January 2022
© The Author(s) 2021

Abstract

Online stores assist customers in buying the desired products online. Great competition in the e-commerce sector necessitates technology development. Many e-commerce systems not only present products but also offer similar products to increase online customer interest. Due to high product variety, analyzing products sold together similar to a recommendation system is a must. This study methodologically improves the traditional association rule mining (ARM) method by adding fuzzy set theory. Besides, it extends the ARM by considering not only items sold but also sales amounts. Fuzzy association rule mining (FARM) with the Apriori algorithm can catch the customers' choice from historical transaction data. It discovers fuzzy association rules from an e-commerce company to display similar products to customers according to their needs in amount. The experimental result shows that the proposed FARM approach produces much information about e-commerce sales for decision-makers. Furthermore, the FARM method eliminates some traditional rules considering their sales amount and can produce some rules different from ARM.

Keywords E-commerce · Recommendation system · Fuzzy rules · Association rule mining · Fuzzy set theory

Introduction

Electronic commerce (e-commerce) has gained increasing popularity, because the Internet becomes an essential tool for companies to increase their competitive edge by collecting and analyzing customer data. The most popular sort of e-commerce is Business-to-Consumer (B2C) sales, which is fulfilled by online stores. The total e-commerce share of total retail sales has increased about 2% each year since 2015, and it is about 20% in 2021 [1]. Retail e-commerce sales increased by more than 20% worldwide in a year, from 2017 (2382 billion USD) to 2018 (2928 billion USD). It is expected that sales volume will reach 6542 billion USD in 2023. Besides, the number of online customers raised about

11% in 2 years, from 2019 (1.92 million users) to 2021 (2.14 million users) [2].

This rising market share is a significant interest for B2C companies to facilitate customers' purchase decision. Moreover, the increase of the digital economy causes heavier competition among companies in the similar industry. On the other hand, exposing many items to customers is one of the main challenges to display appropriate products. For this reason, customers are mainly confused when deciding on a product to buy. Consequently, e-commerce companies have been forced into displaying the right products to engage online customers' attention by supporting their selection decisions. In e-commerce trends, today's research demands involve analyzing customers' online shopping behaviours for their purchasing decision by B2C companies [3]. With these research outcomes, online stores can sharpen their e-commerce platforms and offer a better purchasing experience. One way of presenting practical online user experiences by a well-thought-of website is to associate products [4].

Association Rule Mining (ARM) is a technique to associate products considering their sales frequencies. ARM utilizes customers' transaction data to investigate their demands and product choices [5]. It is powerful to detect relationships between products, which meet specific support and

✉ Onur Dogan
onur.dogan@bakircay.edu.tr

¹ Department of Industrial Engineering, Izmir Bakircay University, 35665 Izmir, Turkey

² Research Center for Data Analytics and Spatial Data Modeling (RC-DAS), Izmir Bakircay University, 35665 Izmir, Turkey

³ Department of Industrial Engineering, Istanbul Technical University, 34367 Istanbul, Turkey

confidence degrees. The discovered association rules are invaluable for B2C companies to offer products that a customer is interested in [6]. Association rules are created by counting products that are purchased together and trust of product relationships, respectively, called support and confidence. These two parameters are critical for generating association rules. A very low support value may generate rules by chance and also cause to produce uninteresting rules. An association rule is regarded as interesting when the support value is greater than the minimum support threshold, and the confidence value is greater than the minimum confidence threshold [7].

If k represents the number of products that will be analyzed, k -items-dataset can produce $2^k - 1$ possible frequent items and $R = 3^k - 2^{k+1} + 1$ possible association rules. However, less than 20% of the generated rules are meaningful for 20% of minimum support rate and 50% of minimum confidence rate. Consequently, most of the computations become wasted [8]. Considering real-life implementations include a huge k value, studies needs a suitable algorithm to avoid this problem. One way to reduce the computational time and cost is to decrease possible frequent items. The second way is to lessen the number of combinations. Apriori algorithm overcomes the problem of a high number of data attributes. It diminishes the number of possible items to produce frequent association rules. According to the Apriori principle, an item set that is not frequent cannot be included in the combined item sets. In other words, all subsets of a frequent item set must also be frequent [7]. It is also known as support-based pruning because of this principle.

The big data era has occurred a vast amount of data. Therefore, many data between datasets overlap, and datasets' boundaries are more ambiguous than in the past. Boolean association rules can handle Boolean data, which is not suitable for overlapping and boundary data points. On the other hand, the fuzzy association rules can generate more reliable solutions even if data overlap and are located in boundaries. If we can obtain fuzzy association rules, we can ideally generate solutions for the problem of product associations. Fuzzy Association Rule Mining (FARM) is an extension of the ARM method by fuzzy set theory. In many real-life datasets, one product may belong to more than one class [9]. For example, a T-shirt can be put under sportswear class and clothing class. Traditional methods cannot consider these multiclass cases. A fuzzy extension of the traditional ARM method can close this gap [10]. Recent studies showed how fuzzy set theory could improve in evaluating the online customers' demand from a B2C company [6,11–13].

This study aims to present the potential of using the FARM approach with Apriori algorithm for e-commerce sales. Sales volumes for each product are the primary attribute for this research. The main contribution of the study is to recommend more relevant products to online customers by considering

not only items sold but also sales amounts. Previous studies such as [1,14] applied in e-commerce focused on items sold together in a certain period. However, the amount of the sold items should be considered to produce more meaningful results by association rule mining. The results of the study support decision-makers deciding how many and which kinds of products to generate the closest attention from online customers. This research proposes a methodology extending association rule mining with fuzzy set theory for this motivation. Although there are various fuzzy-based rule mining approaches in the literature, handling sales amounts distinguishes this study from others. Moreover, this study analyzes products specifically for Muslim women, because the company sells products that meet modest women's desires. To the best of the authors' knowledge, the fuzzy-based rule mining methodology is applied to Islamic products for the first time.

The remaining part of the paper is as follows: "Literature review" presents related works on association rules. "Preliminaries and methodology" explains the proposed FARM methodology. "Evaluation of the proposed methodology" shows the real-world numerical application of the FARM approach. "Conclusion and limitations" concludes the study by discussing the results and giving the limitations of the proposed methodology.

Literature review

E-commerce platforms produce a significant part of a company's income, since they provide personalized recommendations for online customers. The purchasing decisions, experiences, and motivation of customers may change their preference, purchasing process, and final decision [15–21]. To increase people eagerness for online shopping, recommendation mechanisms and personalized systems are broadly utilized on current e-commerce websites [17,22–26]. Several data mining techniques can be formulated to examine the data in e-commerce sales and recommend related products to users. Recommendation systems are applied originally to e-commerce websites for offering products to customers.

Selecting the recommendation method plays a vital role to generate recommendations. Association rule mining can enable recommendations, and Apriori is a well-known ARM algorithm. Reference [27] used clickstream data to develop a recommender system using association rule mining. The developed methodology measures the confidence degrees between clicked products, between the products placed in the basket, and between purchased products, respectively. Reference [28] proposed a personalized recommender system employing ARM and classification. Buyer demands were obtained from text documents and then converted into some

meaningful expressions. Association rules were mined from the transformed phrases using the Apriori algorithm.

The rule-based theory as another common way has been extensively implemented in association rule mining [29]. Fuzzy logic is one of the most popular rule-based methods in the literature. More importantly, some exciting types of research on fuzzy extensions of traditional ARM models were conducted to create effective association rules. Reference [30] created efficient association rules FARM. They improved the fuzzy Apriori algorithm for uncertain data, so that it produces fuzzy association rules. Reference [31] introduced an ARM approach based on fuzzy set theory and demonstrated that fuzzy extension of ARM is more effective. Reference [32] proposed a fuzzy reasoning approach and applied it to assess the created fuzzy association rules. Some recent studies developed the FARM method with machine learning techniques. Reference [31] combined fuzzy association rule mining and fuzzy c-means clustering algorithm to discover association rules. Reference [33] introduced a recommendation system using fuzzy association rules. In this study, a regression model was integrated into the recommendation system of fuzzy association rules to extend the analysis statistically. Combinatorial optimization methods can be combined with recommendation systems. Reference [34] suggested a particle swarm optimization algorithm to recommend multiple products simultaneously.

FARM has been a popular method in recommendation systems in recent years [6]. It has been extended with additional properties to benefit from advantages of fuzziness to new problems such as profile-based fuzzy association rule mining (PB-FARM) approach [35] and class-based fuzzy soft associative (CBFSA) [36]. FARM studies in the literature can be seen in various categories such as classification [36–38], optimization [39], and regression [33].

Reference [37] developed a classifier based on fuzzy association rule-based and applied it for imbalanced classification problems, which include additional challenges for learning methods. Heuristic approaches were combined with the FARM method in some studies. Reference [39] used a differential evolution algorithm to create important fuzzy association rules. Reference [33] proposed a fuzzy recommendation system considering association rules. They analyzed the data statistically and applied the regression model.

Studies based on association rules in the e-commerce environment have focused on various research objectives. In the e-commerce domain, customers exhibit highly changed shopping behaviors [28]. Reference [40] adopted an FP-growth algorithm to analyze orders and develop a batching algorithm for massive customer orders with small batch and high frequency. In large datasets, ARM requires a time-consuming process. Therefore, [28] improved a recommendation system that includes ARM and clustering. They

applied hierarchical clustering techniques to divide a large dataset into a tree of clusters.

Consequently, it can be understood from the above literature review that the FARM method has been applied to various business problems, including overlapping and ambiguous boundaries data. Especially, in e-commerce, the FARM method has been implemented because of much data without a clear boundary. At present, the advancement of the FARM has become a popular research topic [6]. The use of association rules can yield more meaningful information that can be used in the decision-making process.

The reviewed literature indicates that no studies examine the e-commerce product associations purchased by particular customers with an effective method. This study focuses on Muslim women's online shopping transactions with a fuzzy-based association rule mining methodology. A real-life dataset was used to verify the applied methodology.

Preliminaries and methodology

Association rules

An association rule can be expressed as $X \rightarrow Y$, where $X, Y \subseteq I$, $X, Y \neq \emptyset$, and $X \cap Y \neq \emptyset$. I refers a finite set of items and T indicates a finite set of transactions with items in I . The rule $X \rightarrow Y$ explains that every transaction T that includes X contains Y .

The standard criteria to evaluate association rules are support and confidence parameters. Both of them are a subset of items. The support of an item set $I_0 \subseteq I$ is the probability that a transaction T includes I_0 (Eq. 1)

$$\text{supp}(I_0, T) = \frac{|\{\tau \in T \mid I_0 \subseteq \tau\}|}{T}. \quad (1)$$

The support and confidence of an association rule $X \rightarrow Y$ in T is given in Eqs. 2 and 3, respectively. Note that *supp* is used for support value for item set, and *Supp* refers to support value for the association rule

$$\text{Supp}(X \rightarrow Y, T) = \text{supp}(X \cup Y) \quad (2)$$

$$\text{Conf}(X \rightarrow Y, T) = \frac{\text{Supp}(X \rightarrow Y, T)}{\text{supp}(X)} = \frac{\text{supp}(X \cup Y)}{\text{supp}(X)}. \quad (3)$$

The ARM method efforts to mine to create association rules whose support and confidence values are greater than predefined thresholds called *minsupp* and *minconf*, respectively. Such rules are described as strong rules.

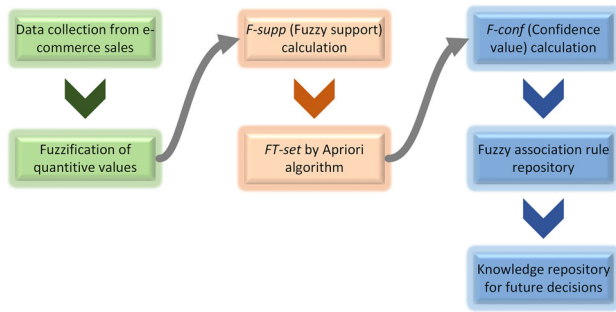


Fig. 1 Proposed FARM methodology

Fuzzy transactions and fuzzy association rules

A fuzzy transaction $\tilde{\tau}$ is nonempty fuzzy subset of item set I . $\tilde{\tau}(i)$ is the membership degree of i in a fuzzy transaction $\tilde{\tau}$ for every $i \in I$. A crisp transaction is a particular form of fuzzy transaction. Similarly, T -set is a set of crisp transactions, whereas FT -set indicates a set of fuzzy transactions. Fuzzy support (F -supp) values of the fuzzy item sets are calculated by the minimum operator given in Eq. 4. The total amount of minimum values of data points t_k in $\tilde{\tau}$ for each product (z_i) is divided into N , which refers to the total number of data points

$$F\text{-supp}(X \rightarrow Y, \tau) = \frac{\sum_{k=1}^N \min\{t_k(z_i)\}}{N} \tag{4}$$

A fuzzy association rule $X \rightarrow Y$ holds if and only if $\tilde{\tau}(X) \leq \tilde{\tau}(Y)$ for each $\tilde{\tau} \in T$. It means that the inclusion degree of Y is greater than that of X for every fuzzy transaction $\tilde{\tau}$. An association rule is a specific form of fuzzy association rule, because a transaction is a specific form of fuzzy transaction. The confidence value of a fuzzy rule is computed by Eq. 5

$$F\text{-conf}(X \rightarrow Y, \tau) = \frac{F\text{-supp}(X \cup Y)}{F\text{-supp}(X)} \tag{5}$$

Proposed FARM model

This study uses a three-stage methodology given in Fig. 1. The first stage given in green is about the database activities. It begins with collecting related sales data from databases by queries. Then, sales amounts are converted into a fuzzy number using predefined fuzzy membership sets. The second stage given in orange is counting frequent products in the fuzzy transactions. First, fuzzy support numbers are counted, and then, the Apriori algorithm runs for all combinations of frequent products to find out frequencies of products sold together. The third stage given in blue discovers fuzzy association rules and stores them in repositories for future usage.

Table 1 Data description

Transaction	Products					
	1075	1078	1079	1084	1086	1087
679		4			12	3
680	1	2	4		3	4
691	7	8	11		2	3
697	11	6		12		
703			10	3		

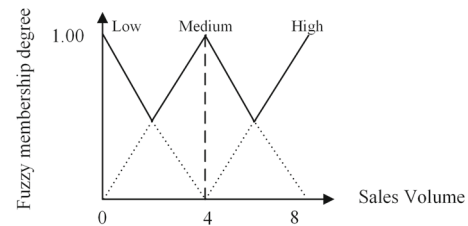


Fig. 2 Fuzzy membership sets for sales amounts

Evaluation of the proposed methodology

A case scenario was presented to illustrate the feasibility of the proposed methodology in this section. The algorithm was applied to 101,250 online transactions including 100 different products. Data were collected from a B2C company, one of the leaders in the sector with more than 16 million visitors from all over the world. The company sells over 500 brands and 70 thousand products to five continents. In this study, special products for Muslim women were selected to verify the proposed FARM methodology.

The first stage of the methodology starts with data preparation steps. Transactions that have only one kind of product were ignored, because this study considers associations among products. Each transaction includes at least two types of the product given in the Products column in Table 1. The proposed FARM methodology considers the sales amount, whereas the traditional ARM method takes into account binary variables, sold or not.

All sales amounts in each transaction were transformed into the fuzzy sets numbers. The fuzzy membership functions in Fig. 2 were determined by company managers considering products’ sales volume. For example, one of the customers purchased three kinds of products in transaction ID 679. The customer bought 4, 12, and 3 items of product 1078, 1086, and 1087, respectively. Although 4 and 12 directly correspond to “Medium” and “High” classes, respectively; 3 lies in both “Low” and “Medium” classes. 3 was transformed into the fuzzy set, calculated as $(0.25/Low + 0.75/Medium)$.

Table 2 presents the converted qualitative numbers into fuzzy sets for each product considering all transactions. After the transformation process, the total fuzzy score (TFS) was

Table 2 Fuzzification of sales amounts

Transaction	Products					
	1075	1078	1079	1084	1086	1087
679		{0.00, 1.00, 0.00}			{0.00, 0.00, 1.00}	{0.25, 0.75, 0.00}
680	{0.75, 0.25, 0.00}	{0.50, 0.50, 0.00}	{0.00, 1.00, 0.00}		{0.25, 0.75, 0.00}	{0.00, 1.00, 0.00}
691	{0.00, 0.25, 0.75}	{0.00, 0.00, 1.00}	{0.00, 0.00, 1.00}		{0.50, 0.50, 0.00}	{0.25, 0.75, 0.00}
697	{0.00, 0.00, 1.00}	{0.00, 0.50, 0.50}		{0.00, 0.00, 1.00}		
703			{0.00, 0.00, 1.00}	{0.25, 0.75, 0.00}		
TFS	{0.75, 0.50, 1.75}	{0.50, 2.00, 1.50}	{0.00, 1.00, 2.00}	{0.25, 0.75, 1.00}	{0.75, 1.25, 1.00}	{0.50, 2.50, 0.00}
FFC	1075H	1078M	1079H	1084H	1086M	1087M

TFS total fuzzy score, FFC final fuzzy class

Table 3 Fuzzy support values of frequent 1-item set

Products	Low	Medium	High	Products	Low	Medium	High
1075L	1082.5	832	75.5	1156L	539.25	343	7.75
1078M	13219.25	15156.5	3468.25	1168L	2293	1813.75	184.25
1079L	3163.25	2222.75	175	1169L	709	404	17
1082L	5133	3363.5	198.5	1171L	954.5	675.75	43.75
1100M	4083.75	4254.75	781.5	1179L	584.5	306.25	5.25
1106L	5882.75	4906.25	409	1180L	911.25	566	31.75
1107L	1131.25	592.75	19	1183L	723	348.75	11.25
1108L	4689.25	3883.75	453	1202L	2387.25	1876.75	215
1110L	5659.5	4067.75	325.75	1232L	2077.75	2011	423.25
1118L	1102	537.25	14.75	1239L	575.75	334.25	9
1119L	2871	1481.75	49.25	1241L	700.75	383.25	11
1120L	510.25	313.25	10.5	1245L	1465	1045.75	72.25
1121M	9859.5	13674.75	4875.75	1246L	1567	927.5	40.5
1126L	3114	2044.5	162.5	1251M	3807	4027	917
1127L	4430.25	3279	296.75	1258L	2095.5	1891.75	244.75
1131L	897	517.5	17.5	1267L	1168.25	798.25	42.5
1140L	1239.25	1087.5	120.25	1269L	1200.25	772.5	38.25
1150M	14871.75	20379	5447.25	1284L	606.75	370.75	18.5

computed by aggregating all fuzzy membership values for each class. The final fuzzy class (FFC) of a product was defined using the maximum TFS. For the previous example, product 1075 has a TFS of {0.75, 0.50, 1.75}. In this case, product 1075 was assigned to the “High” class and renamed as 1075H.

In the second stage, first, fuzzy support values were computed for the 1-itemset of products and given in Table 3. Minimum support (*minsupp*) value was determined by managers as 500 items in all transactions. This means that a product is assigned to the frequent 1-item set if the maximum TFS of the product is larger than *minsupp* value.

According to the Apriori principle, the algorithm eliminates items with a lower support value than the *minsupp* from the frequent item set. The frequent 2-item set was created computing all possible combinations of the frequent

1-item set using the Apriori algorithm. Table 4 shows the fuzzy support values of the 2-item set.

Similarly, the frequent 3-item set was created for all possible combinations of the frequent 2-item set. Table 5 gives the results of frequent item sets that include three products. The algorithm terminated, because none of the 4-item set could meet the *minsupp* threshold.

The third stage creates and stores fuzzy association rules. All potential association rules for each frequent itemsets were extracted, and the *minconf* threshold, 35%, was checked for each created rule. Top 15 fuzzy association rules and the corresponding confidence values are presented in Table 6. For example, the confidence value of the rule “if {1078M, 1121M} then {1150M}” was calculated as

$$\frac{n(1078M \cap 1121M \cap 1150M)}{n(1078M \cap 1121M)} = \frac{3018.75}{8625.00} = 0.35.$$

Table 4 Fuzzy support values of frequent 2-item set

Products	F-supp	Products	F-supp	Products	F-supp
{1078M, 1079L}	602.25	{1079L, 1121M}	587.5	{1121M, 1127L}	856.75
{1078M, 1082L}	1000.25	{1079L, 1150M}	901.75	{1121M, 1150M}	5283
{1078M, 1100M}	1143.5	{1082L, 1110L}	595.75	{1121M, 1232L}	640.75
{1078M, 1106L}	1260.5	{1082L, 1121M}	891.75	{1121M, 1251M}	1201.75
{1078M, 1108L}	1350	{1082L, 1150M}	1364.5	{1126L, 1150M}	552.75
{1078M, 1110L}	1560.25	{1100M, 1121M}	2079.25	{1127L, 1150M}	1326
{1078M, 1121M}	4127.25	{1100M, 1150M}	1431.25	{1150M, 1168L}	847.25
{1078M, 1126L}	812.75	{1106L, 1121M}	1297	{1150M, 1202L}	500.75
{1078M, 1127L}	1235.5	{1106L, 1127L}	556	{1150M, 1232M}	640
{1078M, 1150M}	5047.75	{1106L, 1150M}	2821	{1150M, 1245L}	596
{1078M, 1202L}	860	{1108L, 1121M}	1099	{1150M, 1246L}	547.25
{1078M, 1232L}	530	{1108L, 1150M}	1374.25	{1150M, 1251M}	1391.25
{1078M, 1251M}	976.75	{1110L, 1121M}	1171.25	{1150M, 1258L}	610.75
{1078M, 1258L}	708.25	{1110L, 1150M}	1251.5		
{1079L, 1108M}	770.75	{1121M, 1126L}	579.5		

Table 5 Fuzzy support values of frequent 3-item set

Products	F-supp
{1078M, 1100M, 1121M}	550.25
{1078M, 1106M, 1150M}	711.75
{1078M, 1121M, 1150M}	1417.75
{1078M, 1127L, 1150M}	506.25
{1150M, 1100M, 1121M}	692.75
{1150M, 1121M, 1106M}	772.5

Comparison of FARM, ARM, and discrete ARM

Table 7 is presented to show the contribution of the proposed FARM approach by a comparison with traditional ARM. In both methods, support threshold value and confidence value were determined as 500 and 0.35, respectively.

Traditional ARM discovered created 58 rules, whereas the proposed FARM approach created 22 fuzzy rules. Although ARM presented more rules, FARM included more information in fewer rules, because it gives details about sales volume instead of “sold or not (binary)” information. For example, for one of the common rules that include $X = \{1127 \text{ and } 1150\}$ and $Y = \{1078\}$, ARM indicates that if products 1127 and 1150 sold, then product 1078 is sold. However, FARM shows that if product 1127 is sold over 8 items and 1150 sold up to 8 items, then product 1078 is sold up to eight items. These results confirm that the proposed FARM approach produces much information about e-commerce sales for decision-makers. Furthermore, the FARM method eliminates some traditional rules considering their sales amount and can produce some rules different from ARM, such as “If {1150M, 1100M} then {1121M}” and “If {1078M, 1121M} then {1150M}”.

Table 6 Fuzzy association rules with 35% of confidence value

Fuzzy rules	F-conf	Fuzzy rules	F-conf
If {1121M, 1106M} then {1150M}	0.58	If {1078M, 1127L} then {1150M}	0.40
If {1246L} then {1150M}	0.57	If {1082L} then {1150M}	0.40
If {1245L} then {1150M}	0.56	If {1127L} then {1150M}	0.39
If {1106L} then {1150M}	0.55	If {1121M} then {1150M}	0.39
If {1078M, 1106M} then {1150M}	0.55	If {1126L} then {1078M}	0.38
If {1150M, 1100M} then {1121M}	0.50	If {1127L, 1150M} then {1078M}	0.37
If {1078M, 1100M} then {1121M}	0.50	If {1110L} then {1078M}	0.37
If {1100M} then {1121M}	0.48	If {1258L} then {1078M}	0.37
If {1168L} then {1150M}	0.47	If {1127L} then {1078M}	0.37
If {1202L} then {1078M}	0.45	If {1078M, 1121M} then {1150M}	0.35
If {1079L} then {1150M}	0.40	If {1100M, 1121M} then {1150M}	0.35

Table 7 ARM association rules with 35% of confidence value

Rules	Confidence	Rules	Confidence
If {1121, 1078, 1106} then {1150}	0.64	If {1121, 1127} then {1078}	0.41
If {1100, 1106} then {1150}	0.62	If {1079} then {1150}	0.40
If {1100, 1078, 1150} then {1121}	0.59	If {1078, 1082} then {1150}	0.40
If {1121, 1106} then {1150}	0.58	If {1121, 1110} then {1078}	0.40
If {1246} then {1150}	0.57	If {1126, 1150} then {1078}	0.40
If {1106, 1127} then {1150}	0.57	If {1078, 1127} then {1150}	0.40
If {1106, 1108} then {1150}	0.57	If {1082} then {1150}	0.40
If {1245} then {1150}	0.56	If {1127} then {1150}	0.39
If {1106, 1082} then {1150}	0.56	If {1150, 1258} then {1078}	0.39
If {1106} then {1150}	0.55	If {1121, 1126} then {1078}	0.39
If {1078, 1106} then {1150}	0.55	If {1121} then {1150}	0.39
If {1121, 1168} then {1150}	0.54	If {1140} then {1121}	0.39
If {1121, 1202} then {1078}	0.50	If {1121, 1108} then {1078}	0.38
If {1100, 1150} then {1121}	0.50	If {1232, 1150} then {1121}	0.38
If {1100, 1078} then {1121}	0.50	If {1126} then {1078}	0.38
If {1150, 1202} then {1078}	0.50	If {1078, 1106, 1150} then {1121}	0.38
If {1100} then {1121}	0.48	If {1110, 1150} then {1078}	0.38
If {1168} then {1150}	0.47	If {1100, 1078} then {1150}	0.37
If {1121, 1079} then {1150}	0.47	If {1121, 1251} then {1150}	0.37
If {1131} then {1150}	0.46	If {1127, 1150} then {1078}	0.37
If {1121, 1127} then {1150}	0.46	If {1110} then {1078}	0.37
If {1075} then {1121}	0.46	If {1121, 1232} then {1150}	0.37
If {1202} then {1078}	0.45	If {1258} then {1078}	0.37
If {1107} then {1150}	0.44	If {1127} then {1078}	0.37
If {1100, 1078, 1121} then {1150}	0.44	If {1140} then {1150}	0.36
If {1121, 1082} then {1150}	0.43	If {1078, 1108} then {1150}	0.35
If {1078, 1079} then {1150}	0.43	If {1121, 1078} then {1150}	0.35
If {1171} then {1150}	0.42	If {1100, 1121} then {1150}	0.35
If {1121, 1108} then {1150}	0.42	If {1251, 1078} then {1150}	0.35

Table 8 gives the run time of ARM and FARM, respectively, for different support and confidence threshold values. For example, ARM worked 206.7 s, whereas FARM ran 112.13 seconds for the support threshold is 200 items, and the confidence threshold is 20%. FARM was faster than ARM in all cases except for the support threshold is 600 items, and the confidence threshold is 50%. Textbf cells in the table emphasize the maximum difference between the two methods based on the support values. When the support threshold value of 500 items, determined by managers, was considered, FARM produced results faster for the confidence threshold value of 0.35.

This study also compares the results of discrete ARM and FARM. In discrete ARM, each sales amount is considered Low, Medium, or High. One customer transaction can belong to only one of these classes. FARM also considers

Low, Medium, and High classes, but it calculates membership values for each class.

Discrete ARM requires expanding the product types based on predefined classes. For example, the product 1075 is divided into 1075 low, 1075 medium, and 1075 high sub-products. Therefore, it usually produces fewer association rules than traditional ARM and FARM. However, this study shows that discrete ARM is not a proper method for a large number of products in the e-commerce sector. Table 9 shows the generated number of rules for the corresponding threshold values.

Conclusion and limitations

In this study, some products special for Muslim women were considered to investigate their associations. To do

Table 8 Run time of ARM and FARM (in seconds)

Supp./Conf.	0.2	0.25	0.3	0.35	0.4	0.45	0.5
200	206.7–112.13	191.85–122.98	200.22–116.22	245.48–124.51	198.72–120.07	172.73–121.06	167.88–132.07
300	144.36–67.53	103.6–66.57	91.97–67.23	119.59–59.9	112.34–72.22	91.08–84.88	101.66–85.16
400	127.83–27.64	75.86–43.75	62.94–32.96	124.04–43.3	83.68–33.99	63.89–52.94	61.44–45.64
500	52.68–21.9	55.89–26.06	41.74–25.94	57.59–25.86	53.82–34.91	40.08–39.4	38.99–36.08
600	43.38–22.54	52.57–20.36	30.84–20.37	50.4–24.81	42.64–26.93	31.05–27.68	30.53–35.58
700	52.54–21.37	42.16–17.5	27.2–17.42	29.54–17.5	38.25–21.5	27.41–20.38	27.36–14.97
800	34.07–14.11	39.56–15.51	23.52–16.88	39.66–17.5	37.75–18.11	23.84–16.23	24.98–13.04

Table 9 Comparison table

Methods*	1	2	3	1	2	3	1	2	3	1	2	3
Supp./Conf	0.2			0.25			0.3			0.35		
200	169	466	12	149	404	2	109	320	1	66	227	0
300	101	404	10	90	256	2	63	200	1	37	136	0
400	83	320	10	75	184	2	53	135	1	31	89	0
500	57	227	7	53	132	2	34	95	1	22	58	0
600	46	156	6	43	105	2	29	74	1	17	43	0
700	40	102	4	37	92	2	24	65	1	15	39	0
800	34	67	4	32	83	2	20	58	1	12	36	0
Methods*	1	2	3	1	2	3	1	2	3			
Supp./Conf	0.4			0.45			0.5					
200	40	156	0	28	102	0	19	67	0			
300	18	87	0	12	57	0	7	37	0			
400	14	57	0	10	38	0	7	27	0			
500	11	34	0	10	23	0	7	17	0			
600	8	24	0	7	15	0	4	8	0			
700	7	20	0	6	12	0	3	7	0			
800	5	18	0	4	12	0	1	7	0			

*1: FARM, 2: ARM, 3: Discrete ARM

this, an FARM methodology was proposed and tested in e-commerce sales. Some strong rules among various products were discovered. The results of such research contribute to decision-makers deciding how many and which kinds of products to generate the closest attention from online customers. For example, the rule “if product 1078 and 1121 sold up to 8 items then product 1150 is sold up to 8 items.” This rule shows that the company should recommend product 1150, and the maximum amount of offer should not exceed eight if they want to increase customer attraction. Company may present a special discount for this item and amount.

This study proposes a methodical process of applying the FARM method in the field of e-commerce sales analysis. The proposed methodology presents valuable support to the company managers, because discovered fuzzy rules are straightforward and can simply be adopted to design customer relations. Moreover, the FARM approach generates rules with more information than traditional ARM, because

it considers not only “sold or not (binary)” variables but also sales amount to create fuzzy rules. Managers can have information about which and how many products should be offered to online customers using the proposed approach to increase customer interest.

The proposed methodology was compared to demonstrate its effectiveness with traditional ARM. The results were interesting, because the FARM approach yielded more information with fewer rules. In a common example, ARM indicates that if products 1127 and 1150 are sold, then product 1078 is sold. However, FARM shows that if product 1127 is sold over 8 items and product 1150 is sold up to eight items, then product 1150 is sold up to 8 items. Consequently, the company can offer product 1078 up to 8 items to the customers buying these two products with the corresponding amount. These results confirm that the proposed FARM approach produces much information about e-commerce sales for decision-makers. Furthermore, the

FARM method eliminates some traditional rules considering their sales amount and can produce some rules different from ARM.

There are some important notes for the readers. As the number of transactions increases, support and confidence values should be decreased. Therefore, the support and confidence threshold value is expressed in numbers instead of rates. Additionally, a rule is created by item sets that satisfy the minimum support (*minsupp*) and minimum confidence (*minconf*) values. It does not imply causality. Causality needs cause and consequence of the rule.

For future studies, readers can focus on methodological setup and experimental effects. From the methodological setup perspective, the described fuzzy sets and threshold values can be changed. On the other hand, the dataset can be enlarged with respect to the data attributes, or the same data attributes can be used for various countries.

From the methodological setup perspective, (i) The fuzzy sets used in the study may affect the analysis results. The described fuzzy sets use triangular fuzzy numbers. It is suggested that authors can redesign fuzzy sets with different representation types, such as trapezoidal fuzzy numbers. (ii) Although managers decided on the minimum support and minimum confidence values, authors can investigate the effects of the changes in these two critical thresholds to enhance the reliability of the proposed methodology.

From the experimental effects perspective, (i) This study considers the items and amounts in the transaction data. However, for further research, authors can take into account several e-commerce features such as income, cost, profit, and sales time. The proposed model can be adapted for these features. (ii) Customers in different regions may have different shopping behaviors. Therefore, applying the methodology to different countries and revealing the cross-cultural difference can be an exciting study.

Acknowledgements This work has been financially supported by TUBITAK (The Scientific and Technological Research Council of Turkey), Project Number: 3180641.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as

long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Fatoni C, Utami E, Wibowo F (2018) Online store product recommendation system uses apriori method. *J Phys* 1140:012034 (**IOP Publishing**)
2. eMarketer, Number of digital buyers worldwide from 2014 to 2021 (in billions), <https://www.statista.com/statistics/251666/number-of-digital-buyers-worldwide/> (2021)
3. Wu I-C, Yu H-K (2020) Sequential analysis and clustering to investigate users' online shopping behaviors based on need-states. *Inf Process Manag* 57(6):102323
4. Bleier A, Harmeling CM, Palmatier RW (2019) Creating effective online customer experiences. *J Mark* 83(2):98–119
5. Feng F, Cho J, Pedrycz W, Fujita H, Herawan T (2016) Soft set based association rule mining. *Knowl-Based Syst* 111:268–282
6. Xiang D, Zhang Z (2020) Cross-border e-commerce personalized recommendation based on fuzzy association specifications combined with complex preference model. *Math Prob Eng*
7. Dogan O, Gurcan OF, Oztaysi B, Gokdere U (2019) Analysis of frequent visitor patterns in a shopping mall. In: *Industrial engineering in the big data era*. Springer, Berlin, pp 217–227
8. Tan P-N, Steinbach M, Kumar V (2016) *Introduction to data mining*. Pearson Education India
9. Delgado M, Marín N, Sánchez D, Vila M-A (2003) Fuzzy association rules: general model and applications. *IEEE Trans Fuzzy Syst* 11(2):214–225
10. Öztaysi B, Onar SÇ (2013) Defining the factors that effect user interest on social network news feeds via fuzzy association rule mining: the case of sports news. In: *Data mining in dynamic social networks and fuzzy systems*, IGI Global, pp 334–345
11. Behera RK, Gunasekaran A, Gupta S, Kamboj S, Bala PK (2020) Personalized digital marketing recommender engine. *J Retail Consum Serv* 53:101799
12. Abhirami K (2020) Web usage mining using fuzzy association rule. In: *2016 International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS)*, IEEE, pp 1–4
13. Castro-Schez JJ, Miguel R, Vallejo D, López-López LM (2011) A highly adaptive recommender system based on fuzzy logic for b2c e-commerce portals. *Expert Syst Appl* 38(3):2441–2454
14. Prajapati DJ, Garg S, Chauhan N (2017) Interesting association rule mining with consistent and inconsistent rule detection from big sales data in distributed environment. *Future Comput Inform J* 2(1):19–30
15. Dogan O, Oztaysi B, Fernandez-Llatas C (2020) Process-centric customer analytics: understanding visit purposes of predicted age groups with discovered paths. *J Multiple-Valued Logic Soft Comput* 35(1/2):147–165
16. Dogan O, Fernandez-Llatas C, Oztaysi B (2019) Process mining application for analysis of customer's different visits in a shopping mall. In: *International Conference on Intelligent and Fuzzy Systems*. Springer, pp 151–159
17. Pappas IO, Kourouthanassis PE, Giannakos MN, Lekakos G (2017) The interplay of online shopping motivations and experiential fac-

- tors on personalized e-commerce: a complexity theory approach. *Telematics Inform* 34(5):730–742
18. Wenjun Z, Jun W, Shanchuan X (2015) The probing of e-commerce user need states by page cluster analysis—an empirical study on women's clothes from taobao. com. *Data Anal Knowl Discover* 31(3):67–74
 19. Rohm AJ, Swaminathan V (2004) A typology of online shoppers based on shopping motivations. *J Bus Res* 57(7):748–757
 20. Kau AK, Tang YE, Ghose S (2003) Typology of online shoppers. *J Consum Mark* 20(2):139–156
 21. Moe WW (2003) Buying, searching, or browsing: differentiating between online shoppers using in-store navigational clickstream. *J Consum Psychol* 13(1–2):29–39
 22. Richards J, Li M (2018) The Chinese e-commerce search advertising business: a case study of taobao. *Contemp Manag Res* 14(2):121–142
 23. Pappas IO, Kourouthanassis PE, Giannakos MN, Chrissikopoulos V (2016) Explaining online shopping behavior with fsqca: the role of cognitive and affective perceptions. *J Bus Res* 69(2):794–803
 24. Li Y-M, Wu C-T, Lai C-Y (2013) A social recommender mechanism for e-commerce: combining similarity, trust, and relationship. *Decis Support Syst* 55(3):740–752
 25. Huang Z, Zeng D, Chen H (2007) A comparison of collaborative-filtering recommendation algorithms for e-commerce. *IEEE Intell Syst* 22(5):68–78
 26. Schafer JB, Konstan JA, Riedl J (2001) E-commerce recommendation applications. *Data Min Knowl Disc* 5(1):115–153
 27. Kim YS, Yum B-J (2011) Recommender system based on click stream data using association rule mining. *Expert Syst Appl* 38(10):13320–13327
 28. Zhang X-Z (2007) Building personalized recommendation system in e-commerce using association rule-based mining and classification. In: 2007 International conference on machine learning and cybernetics. Vol. 7, IEEE, pp 4113–4118
 29. Ho GT, Ip W, Wu C-H, Tse YK (2012) Using a fuzzy association rule mining approach to identify the financial data association. *Expert Syst Appl* 39(10):9054–9063
 30. Feng S, Zhang H, Cao J, Yao Y (2019) Merging user social network into the random walk model for better group recommendation. *Appl Intell* 49(6):2046–2058
 31. Zhang C, Li T, Ren Z, Hu Z, Ji Y (2019) Taxonomy-aware collaborative denoising autoencoder for personalized recommendation. *Appl Intell* 49(6):2101–2118
 32. Xiao T, Shen H (2019) Neural variational matrix factorization for collaborative filtering in recommendation systems. *Appl Intell* 49(10):3558–3569
 33. Zhang P, Zhang Z, Tian T, Wang Y (2019) Collaborative filtering recommendation algorithm integrating time windows and rating predictions. *Appl Intell* 49(8):3146–3157
 34. Yera R, Martínez L (2017) A recommendation approach for programming online judges supported by data preprocessing techniques. *Appl Intell* 47(2):277–290
 35. Yavari A, Rajabzadeh A, Abdali-Mohammadi F (2021) Profile-based assessment of diseases affective factors using fuzzy association rule mining approach: a case study in heart diseases. *J Biomed Inform* 116:103695
 36. Rohidin D, Samsudin NA, Deris MM, Association rules of fuzzy soft set based classification for text classification problem. *J King Saud Univ-Comput Inform Sci*
 37. Sanz J, Sesma-Sara M, Bustince H (2021) A fuzzy association rule-based classifier for imbalanced classification problems. *Inf Sci* 577:265–279
 38. Sanz JA, Bustince H (2021) A wrapper methodology to learn interval-valued fuzzy rule-based classification systems. *Appl Soft Comput* 104:107249
 39. Zhang A, Shi W (2020) Mining significant fuzzy association rules with differential evolution algorithm. *Appl Soft Comput* 97:105518
 40. Li Y, Li Y (2018) E-commerce order batching algorithm based on association rule mining in the era of big data. In: 2018 Chinese Control And Decision Conference (CCDC), IEEE, pp 1934–1939

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.