ASSOCIATION RULES MINING FROM IOT RETAIL FUZZY TEMPORAL DATASET

Abstract

The retail industry is being significantly impacted by the two key technologies in the sector, IoT and big data. Businesses now have a wealth of new opportunities to get to know their customers and provide them with specific customer journeys that include product recommendations and tailored experiences based on their past preferences. Finding all potential connections or correlations that hold between the items while taking into account a min_sup and a min_conf is known as mining association rules from a retail dataset. It has been divided into two parts, namely, extraction of frequent sets and determination of the rules that apply to the items in the frequent sets. The "temporal association rules mining" refers to the process of identifying associations between items that hold only within specific time windows and not across the course of the entire dataset. This is an extension of "traditional association rules mining". It involves applying association rules to the items that belong to frequent sets and locating itemsets that are frequent at particular time intervals. The time of the transaction is considered to be non-fuzzy in the aforementioned problem. However, we refer to these fuzzy temporal datasets as situations in which the transaction time is ambiguous. In such datasets as the hour of exchange is fuzzy or loose, we perhaps observed that the arrangement of things is successive in specific fuzzy time spans. Such incessant itemsets are named as locally frequent itemsets throughout fuzzy time spans and the relating association rules as neighborhood association rule throughout fuzzy time stretches. These frequent itemsets and their affiliation rules can't be found in the standard manner due

Authors

M. Shenify
Department of Computer Science
College of Computer Science and IT
AlBaha University
Albaha, Saudi Arabia.
maalshenify@bu.edu.sa

Fokrul A. Mazarbhuiya
Department of Mathematics
School of Fundamental and Applied Sciences
Assam Don Bosco University
Assam, India.
fokrul.mazarbhuiya@dbuniversity.ac.in,

Yussuf Perwej
Department of Computer Science & Engineering
Goel Institute of Technology and Management
Lucknow, India.
yusufperwej@gmail.com

to fluffiness engaged with worldly characteristics. The rules may have a periodicity of some kind. Over a fuzzy time period, such rules are referred to as periodic association rules. We propose here calculation which can extricate locally frequent sets over fuzzy fleeting information and afterward intermittent regular sets/affiliation rules from fuzzy worldly. information and afterward intermittent regular sets/affiliation rules from fuzzy worldly information. Last but not least, we have demonstrated the algorithm's effectiveness by using an example.

Keywords: IoT Technology, Big Retail Data, frequent itemsets, membership function, α-cut.

I. INTRODUCTION

Association rules mining issue was characterized at first [1] by R. Agarwal et al in the application of super-market datasets. Large grocery stores have colossal assortments of records of day to day deals. On the off chance that the purchasing behaviors of the purchasers can be dissected, it will be useful in choices making, for example, what to put marked down, how to put the materials on the racks, how to anticipate future buy and so on. An important data-mining problem has been referred to as locating association rules between items in temporal datasets. The market bin exchange is an illustration of transient dataset.

One of the research fields with the fastest current growth is the Internet of Things. Housing and community services, where there is potential for accident reduction, general efficiency improvement, focus on transparency, service personalization, and end-user payment, are two areas that could benefit from the introduction of this technology. The creation and testing of a predictive association rules-based method for anticipating the necessity of repairs of different components inside the smart home, such as heating, ventilation, and air conditioning, was the focus of the authors' [2] article.

Identification of IoT device usage patterns is crucial in the situation of the smart home. Discovering these patterns and employing to guide decision-making can make carrying out daily tasks simple, comfortable, practical, and autonomous. Given the strict storage and processing limitations of IoT devices, execution information extraction in a distributed method is a computational problem. [3] described a technique for extracting implicit correlations between IoT device behavior through integrated associative analysis. The suggested technique determines the utmost pertinent correlations amongst pairs of activities of various IoT devices based on support, confidence, and lift criteria and recommends their integration via HTTP requests.

As a orientation to finding diverse item deals as a systematic promo to provide to clients based on item frequency of purchase, authors in [4] attempted to increase the effectiveness of the Apriori method for mining of the association rules. The study demonstrated that the modified Apriori Algorithm, when used to implement association rule mining, produces results at a higher performance rate or a shorter runtime rate than the original Apriori Algorithm, and it assisted the organization in choosing customer product deals.

In this chapter, we take into consideration datasets that are imprecise, or fuzzy, temporal. In these datasets, the time at which a transaction occurred is attached to the transactions and is either imprecise or approximate. Due to the nature of temporal features, it may be impossible to determine unknown patterns or connections between items in such large volumes of data. Additionally, some association rules may hold throughout the dataset but only for a small, fuzzy time period. To find such rules it is required to identify itemsets frequent at specific time span, that would be clearly fuzzy on the grounds that the transaction time is always fuzzy. The locally frequent itemset over a fuzzy time interval is the name given to such frequent itemsets [5]. Items can be linked to one another using these locally frequent sets. It is important to note that relations of this kind typically occur on a periodic basis. If the aforesaid locally frequent sets too possess the property of being frequent over certain fuzzy time intervals, i.e. they are referred to as periodic frequent sets if they have periodicity, and the rules that apply to them are referred to as periodic rules over fuzzy time intervals. Using an example, this chapter also demonstrates that the algorithm produces the desired outcome. It is assumed in this chapter that the time stamps are fuzzy with similar membership functions; however, the same method can be suitably modified for datasets with different fuzzy membership.

The section is coordinated as follows: The terms and notations used in this chapter are briefly discussed in section 2. In section 3, we depict the proposed calculation. The consequence of manual experiment with a small dataset is provided in the same segment. In section 4, an experimental study is made with a real-life dataset. In section 5, the periodicity of association rule throughout fuzzy time duration is discussed. In section 6, we briefly mention some real-life applications. Last but not least in section 7, a conclusion and lines for subsequent work is given.

II. TERMS, NOTATIONS AND SYMBOLS USED

Let X represents the universe of discourse. A membership function F(a) lying in [0,1] defines a fuzzy set F in X. F(a) for $a \in X$ denotes the membership grade of $a \in F$. Consequently, a fuzzy set F is characterized by

$F=\{(a, F(a)), a \in X \}$

If F(a) = 1 for at least one $a \in X$, then a fuzzy set F is said to be normal.
An α-cut F_α [6] of a fuzzy set A is expressed as

$F_\alpha=\{a \in X; F(a) \geq \alpha\}$

If all of a fuzzy set's **α**-cuts are convex sets, the fuzzy set is said to be convex [7].
A convex normalized fuzzy set A defined on the real line R is a fuzzy number if and only if

- There is an $a_0 \in R$ for which $F(a_0) =1$,
- F(a) is piecewise continuous.

A fuzzy number is denoted by [x, y, z] with $x < y < z$ where F(x) = F(z) = 0 and F(y) = 1. F(a) for all $a \in [x, y]$ is known as left reference function and F(a) for $a \in [y, z]$ is known as the right reference function. The α-cut of the fuzzy number $[f_1-x, f_1, f_1+x]$ is a closed interval $[f_1+(\alpha-1).x, f_1+(1-\alpha).x]$.

Special fuzzy numbers that satisfy the following are called fuzzy intervals.

- there is an interval $[y, z] \subset R$ in which $F(a_0) =1 \ \forall \ a_0 \in [y, z]$,
- F(a) is piecewise continuous.

A fuzzy interval A is a fuzzy number with having a flat region and is symbolized by F = [x, y, z, t] with $x < y < z < t$ where F(x) = F(t) = 0 and F(a) = 1 $\forall \ a \in [y, z]$. F(a) for $a \in [x, y]$ is known as LR function and F(a) for $a \in [z, t]$ is known as the RR function [8].

Accordingly, the α-cut of $[f_1-x, f_1, f_2, f_2+x]$ is $[f_1+(\alpha-1).x, f_2+(1-\alpha).x]$, a closed interval.
Also core of a fuzzy number F is given by, Core(F) = {(a, F(a); F(a) = 1}

For each F, $F = \bigcup_{\alpha \in [0,1]} F_\alpha$

Let F and G be two fuzzy sets then $\forall\ \alpha\in[0, 1]$, the following results hold.

- $(F\cup G)_\alpha = F_\alpha \cup G_\alpha$
- $(F\cap G)_\alpha = A_\alpha \cap B_\alpha$

Let $\Gamma = [\tau_0, \tau_1,...]$ be a list of ambiguous or fuzzy time stamps over which a linear ordering is specified, where $\tau_i < \tau_j$ indicates that τ_i denotes the core of a fuzzy time stamp that is earlier than τ_j. We make the convenient assumption that the membership functions of all fuzzy time stamps are equivalent. Let's say *I* referring to a limited number of items and the transaction database. Each transaction in the collection *D* has a portion that is a subset of the itemset *I* and a portion that is a fuzzy time-stamp giving an approximation of the time that the transaction took place. *D* is assumed to be arranged in the same ascending order as the fuzzy time stamp core. A transaction is considered to be in the fuzzy time interval $[\tau_1-x, \tau_1, \tau_2, \tau_2+x]$ if the α-cut of the fuzzy time stamp of the transaction is contained in the α-cut of $[\tau_1-x, \tau_1, \tau_2, \tau_2+x]$ for some user's specified value of α. For fuzzy time intervals, we always take into account a fuzzy closed interval of the form $[\tau_1-x\ \tau_1, \tau_2, \tau_2+x]$.

We characterize the local support of an itemset in a fuzzy time span $[\tau_1-x, \tau_1, \tau_2, \tau_2+x]$ as the proportion of the quantity of transactions in the time interval $[\tau_1+(\alpha-1).x, \tau_2+(1-\alpha).x]$ containing the itemset to the all-out number of transactions in $[\tau_1+(\alpha-1).x, \tau_2+(1-\alpha).x]$ for the entire *D* for a given worth of α. Given an α, we say that an itemset *F* is frequent in the fuzzy time interval $[\tau_1-x\ \tau_1, \tau_2, \tau_2+x]$ if $supp(F) \geq \alpha$. We say that an association rule $F\rightarrow G$ holds in the time interval $[\tau_1-x\ \tau_1, \tau_2, \tau_2+x]$ iff the given threshold is met. *F* and *G* are the item sets, and $F\cup G$ is frequent in $[\tau_1-x\ \tau_1, \tau_2, \tau_2+x]$.

There is a list of fuzzy time intervals where every locally frequent item set occurs. [*start*-*x*, *start*, *end*, *end*+*x*] denotes each fuzzy interval, with start denoting the approximate start time and end denoting the approximate end time. end - start is the length of the core. If their α-cuts are non-overlapping, the two intervals [*start*1-*x*, *start*1, *end*1, *end*1+*x*] and [*start*2-*x*, *start*2, *end*2, *end*2+*x*] are non-overlapping for a given value of α.

III. METHOD PROPOSED

A sequence of fuzzy time-intervals where the set is frequent is saved for each locally frequent set when it is generated. This is accomplished by utilizing the two user-set predetermined limits α and *minthd*. The present transaction is added in the present time-span, that is reached out by supplanting Rlastseen by Rcurrent, if for a certain itemet the α-cut of its present fuzzy time-stamp, [Lcurrent, Rcurrent], and its fuzzy time's α-cut, [Llastseen, Rlastseen], while it was most recently seen, overlaps during the technique's pass through the dataset. If not, Lcurrent is used as the starting point for a new time interval. The *support* for the itemset is verified to see if it was frequent in the previous time period. Assuming this is the case, fuzzified and added to the rundown is kept up for the set. Eventually, for the locally frequent sets over a fuzzy time interval, user-specified core length is given as *minthd* and fuzzy time period length more than or equal to *minthd* are kept.

While searching the dataset for each item, we always keep a $lastseen_\alpha$=[Llastseen, Rlastseen] and indicates the fuzzy time stamp at which the item was last seen. At the point when a thing is found in a transaction and the fuzzy time-stamp is *tm* and on the off chance that its α-cut has non-overlapping with [$Llastseen_{\alpha}$, $Rlastseen_{\alpha}$], a new interval is begun by putting the new time interval's start as Ltm_{α} and finish of the past time interval as $Rlastseen_{\alpha}$. If the item's support in the previous interval is greater than *min-sup*, it is fuzzified. The fuzzified interval is then added to the list kept with for that itemset.

Two supports are retained, itemcount and transcount. A set is said to be a locally frequent over a fuzzy time interval if its support counts in α-cut is greater than or equal to user-specified threshold *min_sup*.

The above will give every 1-sized locally itemsets along fuzzy time interval lists Then Apriori candidate generation is utilized to find candidate of size 2. We associate a list of fuzzy time intervals that were obtained during the pruning phase with each 2-sized candidate frequent set. In the generation stage this list is empty. A candidate set is constructed if all of its subsets are found at the previous level. The procedure involves selecting the fuzzy interval-list linked with the set from the list of the first subset that is found in the level before it. At the point when resulting subsets exist then the procedure is recreated by considering all conceivable pair-wise non-empty subsets one from each list. This list is further pruned for empty sets.

To outline the aforesaid strategy, a two-days dataset comprising of transactions and fuzzy time stamps. When a transaction is given a fuzzy time stamp, it indicates that the transaction occurred in the fuzzy time. All of the fuzzy time stamps are represented by triangular fuzzy numbers for ease of use. The dataset is given beneath:
Set of transactions with uncertain time stamps.

[0, 2, 4]	1 3 4 5
[1, 3, 5]	1 4 6 7 9
[2, 4, 6]	2 3 5 6 7 10
[3, 5, 7]	1 3 5 7 10
[4, 6, 8]	2 3 4 8 9
[5, 7, 9]	1 2 4 5 6 10
[6, 8, 10]	1 2 3 4 7 8 10
[7, 9, 11]	1 2 3 4 5 6 7 8
[8, 10, 12]	1 2 3 4 5 7 8 9
[9, 11, 13]	2 3 4 6 7 10
[10, 12, 14]	1 2 3 4 5 7 10
[11, 13, 15]	1 2 8 10
[12, 14, 16]	1 2 3 4 5 7
[13, 15, 17]	2 3 6 7
[14, 16, 18]	1 2 3 4 5
[15, 17, 19]	1 2 3 4
[16, 18, 20]	2 5 6 7
[17, 19, 21]	1 2 3 4 5
[18, 20, 22]	2 4 8 10
[19, 21, 23]	2 3 6 9
[20, 22, 24]	3 4 7 8
[21, 23, 1']	1 6 10

[22, 24, 2’]	1 5
[23, 1’, 3’]	7
[24, 2’, 4’]	2 10
[1’, 3’, 5’]	3 4
[2’, 4’, 6’]	1 4 5 8 9 10
[3’, 5’, 7’]	1 2 4
[4’, 6’, 8’]	2 4 5
[5’, 7’, 9’]	2 3 4 6 7
[6’, 8’, 10’]	3 5
[7’, 9’, 11’]	1 3 4 5 6 7
[8’, 10’, 12’]	2 5
[9’, 11’, 13’]	1 2 3 7 8
[10’, 12’, 14’]	1 9 10
[11’, 13’, 15’]	1 2 3 6 10
[12’, 14’, 16’]	2 3 4
[13’, 15’, 17’]	1 2 3
[14’, 16’, 18’]	1 2 3
[15’, 17’, 19’]	3 4 5 7 8
[16’, 18’, 20’]	1 2 3 8 10
[17’, 19’, 21’]	2 3 5
[18’, 20’, 22’]	1 2 3 4 5 6
[19’, 21’, 23’]	1 2 3 5 7 8
[20’, 22’, 24’]	2 3 4 5 9 10

With the aforementioned dataset, we manually run the algorithm, using the parameters min_sup = 0.4, minthd = 3. After executing the method, the following frequent itemsets are obtained.

{({1 2 3 4 5 7}; [7, 9,14, 16]), ({2 3 4 8}; [4, 6, 10, 12]), ({1 2}; [9’, 11’, 21’, 23’]), ({1 3}; [7’, 9’, 21’, 23’]), ({1 10}; [3, 5, 8,10]), ({2 3}; [9’, 11’, 22’, 24’]), ({2 5}; [17’, 19’, 22’, 24’]), ({3 5}; [15’, 17’, 22’, 24’]), ({3 7}; [5’, 7’, 11’, 13’]), ({5 10}; [2, 4, 7, 9]), ({4}; [1’, 3’, 9’, 11’]), ({5}; [2’, 4’, 10’, 12’])}

Generating Association Rules: If any itemset is frequent at [τ_1-x, τ_1, τ_2, τ_2+x], all its subsets are also frequent here. However, the subset’s support at [τ_1-x, τ_1, τ_2, τ_2+x] are required to construct the association rules as indicated in section 2. One more dataset scan is required for this which gives association rule valid over the fuzzy interval.

IV. EXPERIMENTAL STUDY

For the sake of experimental, we have used a retail data [9]. The dataset is a collection of supermarket transactions from Belgian supermarket. The dataset is non-temporal. We incorporate the fuzzy time-stamp in it. The time attributes are incorporated in such a way that the transactions are appears to be ordered according the cores fuzzy time-stamps. The time attribute is incorporated in such a that the life time of the dataset appears to be almost one year. For running we, take different values of α say α=0.2, 0.4, 0.6, 0.8 and also different values of *minthd* say *minthd*=5, 10, 15, 20, 25, 30. The programs are executed for the different combinations of above parameters and the results are recorded. The details of the obtained results are presented graphically in Fig. 1and Fig. 2, using bar diagrams in Fig. 3 and Figure 4.

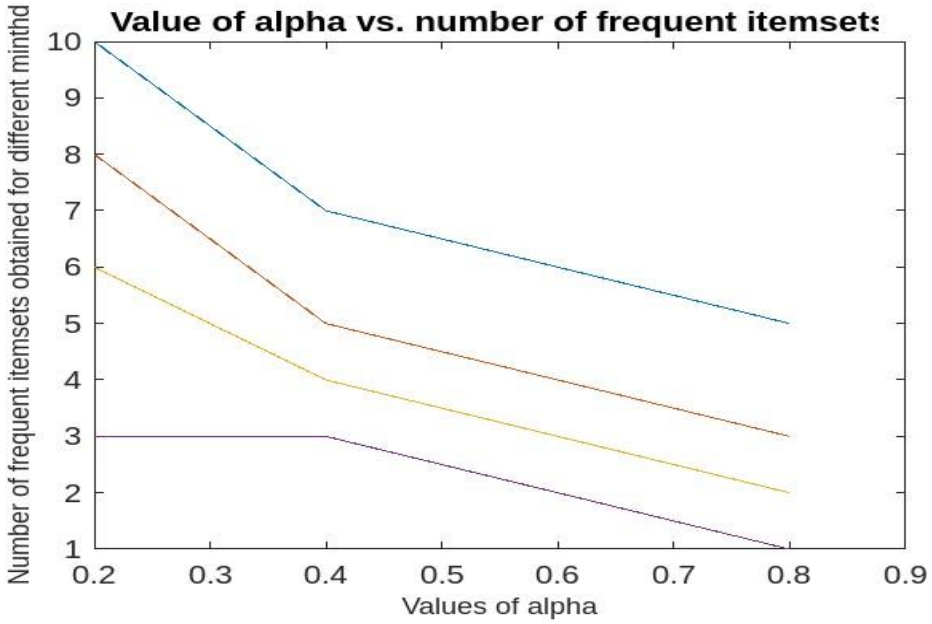

Figure1: The number of frequent itemsets obtained versus α for different minthd

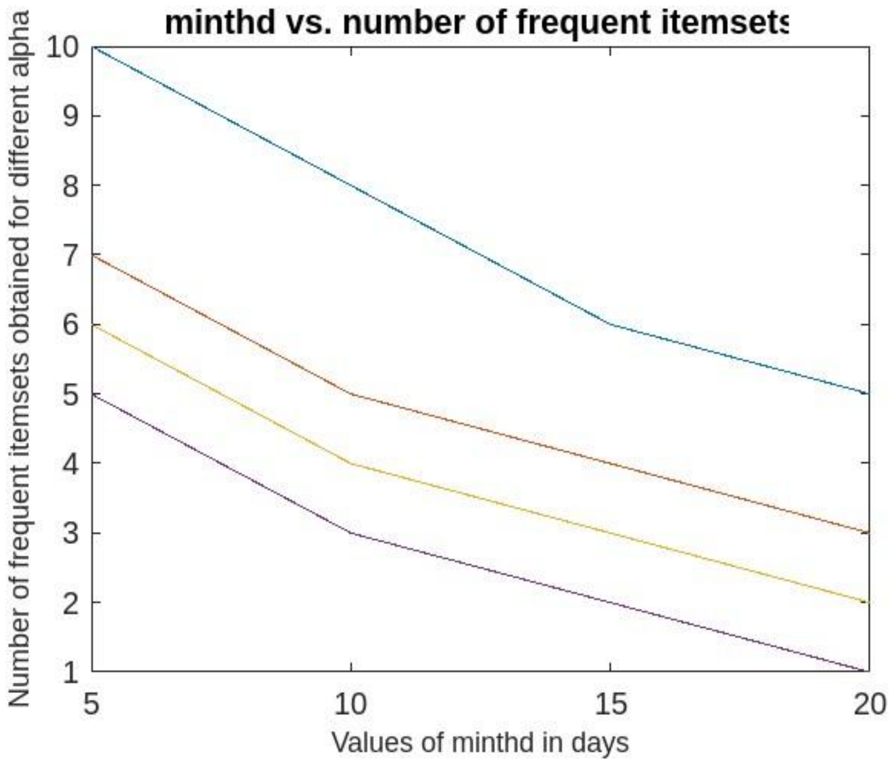

Figure 2: The number of frequent itemsets obtained versus *minthd* for different α.

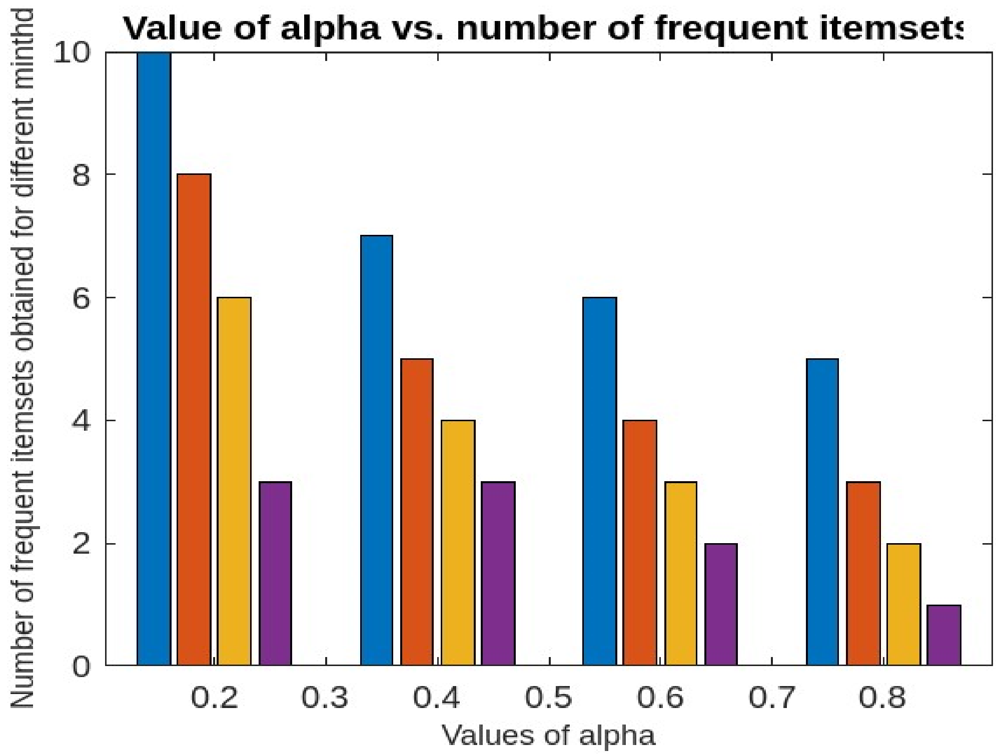

Figure 3: The number of frequent itemsets obtained versus α for different minthd

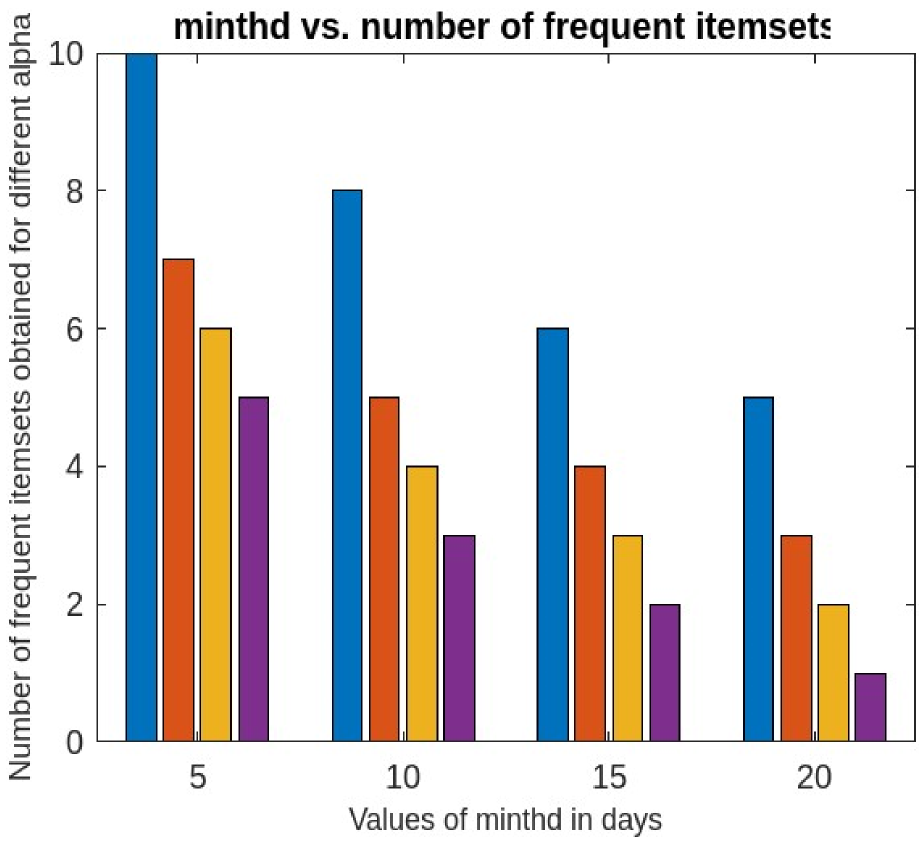

Figure 4: The number of frequent itemsets obtained versus minthd for different α.

It has been observed from the obtained results that the number of frequent itemsets decreases with to increase in the values of either α or minthd or both.

V. EXTRACTING PERIODIC PATTERNS

A fuzzy period list where the set is frequent is kept in the approach suggested in section 3 for each locally frequent set. We refer to them as periodic over fuzzy time intervals if we discover their core-length in the list is nearly same. Periodic association rules over fuzzy time intervals are association rules that are valid on a regular basis.

VI. APPLICATIONS IN IOT

The Association Rule has a variety of uses, including the following:

- Purchases made with a credit card, such as hotel stays and rental vehicles, provide information about the next item that customers are likely to purchase.
- Extra services that teleconnection users choose to pay for (such as call waiting, call forwarding, DSL, speed calls, etc.) let you choose how to combine these features to get the most money.
- Banking services utilized by retail customers (money market accounts, certificates of deposit, investment services, auto loans, etc.) anticipate customers' need for further services.
- An unusual collection of insurance claims might specify fraud which lead to further inquiry.
- The medical patient histories can back up statements about potential problems based on a certain course of therapy.

VII. CONCLUSION AND LINES FOR FUTURE WORK

The study provides an approach for extracting frequent sets from fuzzy temporal data that are frequent in specific fuzzy time periods. The approach computes the frequent sets and the fuzzy time intervals when they occur on a dynamic basis. The aforesaid frequent itemsets are named as locally frequent sets over fuzzy time intervals. The employed method resembles the A priori algorithm. Interesting rules may result from these locally frequent sets. These regionally frequent sets may include certain periodic elements. The associated rules are referred to as periodic association rules over fuzzy time intervals, and these sets are known as periodic frequent sets over fuzzy time intervals.

In the level-wise generation of locally frequent sets, a fuzzy time interval list is maintained where each locally frequent set is frequent. Pair-wise intersection of the intervals in two lists is used to create candidates for the following level. Furthermore, we manually verified the algorithm's effectiveness using an example. For ease of use, we have used a dataset with fuzzy time stamps that have a comparable membership function, but the approach can also be used with datasets that have different fuzzy time stamps. Experimentally, it has been found that the method works well with any datasets having fuzzy time-stamps. Finally, some real-life applications are mentioned briefly.

REFERENCES

[1] R. Agrawal, T. Imielinski and A. Swami; Mining association rules between sets of items in large databases; Proceedings of the ACM SIGMOD '93, Washington, USA (May 1993).
[2] Vasiliy S. Kireev, Anna I. Guseva, Pyotr V. Bochkaryov, Igor A. Kuznetsov & Stanislav A. Filippov, Association Rules Mining for Predictive Analytics in IoT Cloud System, BICA 2018: Biologically Inspired Cognitive Architectures 2018 pp 107–112.
[3] M. Alencar, R. Barreto, H. Fernandes, E. Souto, and R. Pazzi' DARE: A decentralized association rules extraction scheme for embedded data sets in distributed IoT devices, International Journal of Distributed Sensor NetworksVolume 16, Issue 10, October 2020 https://doi.org/10.1177/1550147720962999
[4] Jeanie R. Delos Arcos, Alexander A. Hernandez, Analyzing Online Transaction Data using Association Rule Mining: Misumi Philippines Market Basket Analysis, ICIT '19: Proceedings of the 2019 7th International Conference on Information Technology: IoT and Smart CityDecember 2019Pages 45–49https://doi.org/10.1145/3377170.3377226.
[5] G. J. Klir and B. Yuan; Fuzzy Sets and Fuzzy Logic Theory and Applications, Prentice Hall India Pvt. Ltd. (2002).
[6] G. Q. Chen, S. C. Lee and E. S. H. Yu; Application of fuzzy set theory to economics, In: Wang, P. P., ed., Advances in Fuzzy Sets, Possibility Theory and Applications, Plenum Press, N. Y.(1983), 277-305.
[7] D. Dubois and H. Prade; Ranking fuzzy numbers in the setting of possibility theory, Information Science 30(1983), 183-224.
[8] R. Motwani, E. Cohen, M. Datar, S. Fujiware, A. Gionis, P. Indyk, J. D. Ullman and C. Yang; Finding interesting association rules without support pruning, Proceedings of the16th International Conference on Data Engineering (ICDE), IEEE, (2000).
[9] Tom Brijs, G. Swinnen, K. Vanhoof and G. Wets, "Using association rules for products assortment decisions: A case study", In Knowledge Discovery and Data Mining (1999), pp.254-260.