One-Phase Temporal Fuzzy Utility Mining

Tzung-Pei Hong
Department of Computer Science and Information Engineering
National University of Kaohsiung
Taiwan
tphong@nuk.edu.tw

Cheng-Yu Lin
Department of Computer Science and Engineering
National Sun Yat-sen University
Taiwan
cylin910021@gmail.com

Wei-Ming Huang
Department of Computer Science and Engineering
National Sun Yat-sen University
Taiwan
granthill168@gmail.com

Shu-Min Li
Department of Computer Science and Engineering
National Sun Yat-sen University
Taiwan
smli@cse.nsysu.edu.tw

Shyue-Liang Wang
Department of Computer Science and Information Engineering
National University of Kaohsiung
Taiwan
slwang@nuk.edu.com

Jerry Chun-Wei Lin
Department of Computing, Mathematics, and Physics
Western Norway University of Applied Science
Norway
jerrylin@ieee.org

***Abstract*—Temporal utility mining is complicated than utility mining due to the time consideration. The former discusses the situation that different items may have different profit values and on-shelf time periods. In this paper, we handle the problem of temporal fuzzy utility mining, which simultaneously considers the temporal factor, purchased quantities, item profits and linguistic terms from a transaction database. In the past, we proposed a two-phase tree-based approach to solve this problem. In this paper, we further design a tree structure with less memory to store all the required information and with faster execution speed than the previous one. The experimental results show that the proposed method can get good performance on the execution efficiency and consumed memory size.**

Keywords—data mining, fuzzy set, fuzzy data mining, temporal utility mining, tree structure.

I. INTRODUCTION

Frequent-itemset mining (FIM) uses a user-predefined count threshold called *minsup* to determine the importance of each itemset [1][2][3]. It can help managers to find useful data and make effective decisions. FIM is a popular data mining task in finding association rules or knowledge patterns from a dataset. It assumes all the items in a transaction are equally important, and the mining is conducted according to the appearing frequencies of itemsets in a transaction dataset. The processing strategies of FIM can be classified into level-wise methods and pattern-growth methods. For example, the *Apriori* algorithm is initially proposed to discover knowledge patterns in a level-wise manner. However, more computation and consumed memory may thus occur while the size of the transactional datasets is getting high. Therefore, Han et al. proposed an approach called FP-Growth (Frequent-Pattern Growth) [8] to overcome the problem. FP-Growth uses a tree structure called FP-Tree (Frequent-Pattern Tree) to mine frequent itemsets.

One drawback of association-rule mining is that it does not consider the quantity and profit of an item. Consequently, the utility data mining [21] has recently been viewed as an important data mining task. It attempts to find itemsets with high-utility values based on the quantity information and the profit value of an item. If an itemset is a high utility one, its utility value is larger than or equal to a user-predefined minimum utility threshold. To enhance execution efficiency, some two-phase methods [16][17][22], which use upper-bounds on the utility measure, were proposed to mine high-utility itemsets. Besides, incremental handling for utility mining [20] and weighed applications [5][7][9] were also discussed.

In the past, the concept of the fuzzy set theory [23] has often been used by intelligent systems [24][25][26] because of its simple concept and similar to human thinking. For utility mining, Lan et al. proposed a fuzzy utility mining approach [14] to reflect the factors such as quantity values for items, unit profits, and the implication meaning of quantities. Huang et al. then considered the temporal fuzzy utility mining, which was complicated than fuzzy utility mining due to the time consideration [10]. The former discusses the situation that different items may have different on-shelf time periods. In addition, Hong et al. proposed a two-phase tree-based approach to solve this problem [6]. In this paper, we will introduce another alternative with less memory requirement and less computation time than that in [6]. The proposed approach needs to be executed in only one phase.

II. REVIEW OF RELATED WORKS

FIM [1][2][3], which focuses on discovering information or knowledge patterns in a dataset, is a fundamental data mining task. Its concept is used to facilitate the development of business strategies or decision-making. Note that when a consumer purchases product *A*, he/she may also buy other products in addition to *A*. Agrawal et al. proposed the famous *Apriori* algorithm to discover the relationship of products. Han et al. [8] then designed a compressed tree, called FP-Tree (frequent pattern tree) and proposed the Frequent-Pattern Growth (FP-Growth) algorithm based on it to explore effectively frequent itemsets without candidate itemsets. This FP-Tree was used to store items with relevant information, such as the items and their appearance times. According to FP-Tree, all high frequent

978-1-7281-6932-3/20/$31.00 ©2020 IEEE

itemsets can be found. In the past, some methods [11][15][18][19] based on it have been developed.

Only consider the frequency values of items in a dataset as a factor is not sufficient to reveal the high-profit items in real applications. To overcome this problem, utility mining [4][16][17][21] has been proposed. Utility mining can be considered as an extension of frequent mining, and that information, such as sold quantities and profit values of items, was used to decide the importance of an item or itemset in a dataset. Its goal is to find valuable or high-profit itemsets, but quantity information is not easily comprehended by decision-makers.

Fuzzy utility mining [10][14], which stems from the concept of utility mining and fuzzy data mining [16][17][21][27][28], improved the above problem. Its goal is to consider sold quantitative, profit value, and transformed linguistic terms, which viewed as the implication of quantity to derive itemsets with high fuzzy-utility values. Fuzzy utility mining is, however, needed to perform an exhaustive search on discovering the complete set of high fuzzy-utility itemsets, which caused information loss problem. To resolve this problem, an upper-bound method was proposed to keep the monotonic property in fuzzy utility mining [14]. Based on the extension of the concept as mentioned above, the transaction periods in a dataset then were then considered to discover information patterns with the temporal property of the items [10].

In this paper, we modify our previous two-phase temporal fuzzy utility mining method [6], which applied the upper-bound model [10] to hold monotonic property, to find all possible itemsets in one phase. An extended node structure is designed, which embeds more information (such as transaction ID, fuzzy value and utility value) in addition to the upper-bound value used in our previous method [6], to speed up the mining process.

III. ALGORITHM

A. Problem Statements

A running example in Table 1 is given to illustrate the proposed fuzzy utility mining with the temporal property. The table includes four features and six transactions. A time period can contain one or more transactions, or no transactions and it represents an interval of time. The profits for items are given in Table 2. The membership functions, including *Low*, *Middle*, and *High* fuzzy regions, for all items shown in Figure 1.

TABLE 1. A QUANTITATIVE DATABASE

Period	*TID*	*Items and Quantities*
P_1	$Trans_1$	{6*A*, 2*C*}
	$Trans_2$	{4*A*, 4*B*}
P_2	$Trans_3$	{1*A*, 1*C*}
	$Trans_4$	{3*A*, 3*B*, 4*C*, 4*D*, 2*E*}
P_3	$Trans_5$	{3*A*, 6*B*, 2*E*}
	$Trans_6$	{3*A*, 4*D*, 3*E*}

The previous methods [6][10] were used to solve temporal fuzzy utility mining problem. The corresponding definitions according to those methods are described below.

Definition 1: Let I be the set of all items $\{i_1, i_2, i_3, \ldots, i_m\}$ considered in a dataset. If $X \subseteq I$, then X is an itemset in I. For instance, $\{A, B\}$ is a 2-itemset in I where $I = \{A, B, C, D, E\}$.

TABLE 2. ITEM UTILITIES

Item	*Profit*
A	2
B	6
C	4
D	2
E	4

Definition 2: A time period set $P = \{p_1, p_2, p_3, \ldots, p_t\}$, t is the number of the time periods, and p_t denotes the t-th time period. The time period p_t may contain zero, one, or more transactions.

Definition 3: A temporal quantitative transaction contains four fields: a period of transaction occurrence, the purchased items, the purchased quantities of items, and the item identifications. For instance, take items *A*, *B* and *E* in $Tran_5$ of Table 1: their quantities are 3, 6 and 2, respectively. And they occur in period P_3.

Definition 4: The temporal quantitative transaction database is called $TQD = \{Trans_1, Trans_2, Trans_3, \ldots, Trans_z\}$, where $Trans_i$ is the i-th transaction in TQD and z denotes the number of transactions in TQD.

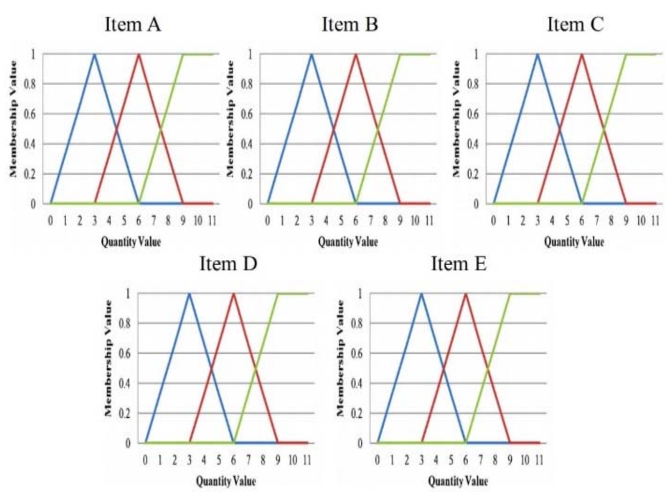

FIGURE 1. MEMBERSHIP FUNCTIONS OF ITEMS

Definition 5: Given the membership functions for every item, we compute the fuzzy values for every transaction. Given an item i_z, its fuzzy set f_{yz} of the value v_{yz} in the temporal quantitative transaction $Trans_y$ is represented as:

$$f_{yz} = \left\langle \frac{f_{yz1}}{R_{z1}} + \frac{f_{yz2}}{R_{z2}} + \cdots + \frac{f_{yzh}}{R_{zh}} \right\rangle.$$

For instance, the quantity value of B in $Trans_2$ can be converted into $\left\langle \frac{0.67}{B.Low}, \frac{0.33}{B.Middle}, \frac{0}{B.High} \right\rangle$ according to Figure 1.

Definition 6: An item i_z has an external utility represented as $s(i_z)$ for the profit of i_z. For instance, $s(A)$ is 2 in Table 2.

Definition 7: The utility value u_{yz} of an item i_z in $Trans_y$ for is considered as

$$u_{yz} = v_{yz} * s(i_z).$$

For example, $u(B)$ in $Trans_2$ is 4*6 (=24).

Definition 8: The fuzzy utility f_{yzl} of its l-th fuzzy region in $Trans_y$ for an item i_z is considered as

$$fu_{yzl} = f_{yzl} * u_{yz}.$$

As mentioned in the above definitions, the fuzzy value of the fuzzy item {*B.Low*} in $Trans_2$ is 0.67. Then, the fuzzy utility value of fuzzy item {*B.Low*} in $Trans_2$ is calculated as 0.67 * 24 = 16.08.

Definition 9: The fuzzy transaction utility tfu_y of a temporal quantitative transaction $Trans_y$ is the sum of the items of the fuzzy utility in $Trans_y$. It is represented as

$$tfu_y = \sum_{i_z \subseteq Trans_y} fu_{yz}.$$

For instance, $tfu_1 = fu_{1,\{A.Middle\}} + fu_{1,\{C.Low\}} = 12 + 5.36 = 17.36$.

Definition 10: The fuzzy utility value fu_{yX} of the fuzzy itemset X in $Trans_y$ is defined

$$fu_{yX} = f_{yX} * \sum_{R_{zl} \subseteq X} v_{yz} * s(i_z),$$

where the value f_{yX} is the membership degree value of X in $Trans_y$, v_{yz} is the quantity of item i_z in X, and $s(i_z)$ is the profit value of item i_z. Value f_{yX} must be the smallest one in R_{zl} in X. For example, the fuzzy itemset {*A.Middle*, *C.Low*} in $Trans_1$ and the fuzzy values are 1 and 0.67. We choose the smallest fuzzy value, 0.67, to calculate the fuzzy utility value of {*A.Middle, C.Low*} in $Trans_1$. The fuzzy value of {*A.Middle, C.Low*} in $Trans_1$ is calculated as 0.67 * [(6 * 2) + (2 * 4)] = 12.8.

Definition 11: The start transaction period STP_{i_z} of an item i_z is represented as the first transaction occurring during the time period of i_z in the *TQD*. In Table 1, the *STP* of *D* is P_2.

Definition 12: The transaction period of itemset X is called LTP_X. For instance, the *STP* of itemset {*A, D*} is P_2 and its last transaction period is the end of the *TQD*, which is P_3.

Definition 13: The temporal fuzzy utility ratio $tfur_X$ of the itemset X is defined as

$$tfur_X = \sum_{X \in Trans_y \cap Trans_y \in LTP_X} fu_{yX} \Big/ \sum_{Trans_y \in LTP_X} tfu_y$$

In the formula, the value fu_{yX} is represented as the *fu* value of the fuzzy itemset X in $Trans_y$ and tfu_y is the transaction fuzzy utility value of $Trans_y$. For instance, the time period of the fuzzy itemset {*B.Low*} in Table 1 is from P_1 to P_3; the first transaction is $Trans_2$ and the last is $Trans_6$. The *tfur* of the fuzzy itemset {*B.Low*} is calculated as (16.08 + 18) / (17.36 + 32 + 1.98 + 53.34 + 47.36 + 26) = 34.08 / 178.04 = 0.19.

Definition 14: Consider X and λ be fuzzy itemset and threshold value, respectively. If $tfur_X >= \lambda$ for any itemsets X, X is a high temporal fuzzy utility itemset (*HTFUI*). For instance, the {*B.Low*} in Table 3 is $HTFUI_1$ if λ is 15%. Note that the monotonic property is not kept in temporal fuzzy utility mining. We must find the candidate itemsets and determine whether these candidate itemsets are *HTFUIs*.

Definition 15: The maximal transaction fuzzy utility of $Trans_y$ is represented as

$$mtfu_y = \sum_{i_z \subseteq Trans_y} mfu_{yz}$$

For instance, in $Trans_2$, $mtfu_2 = mfu_{2,\{A\}} + mfu_{2,\{B\}} = 5.36 + 16.08 = 21.44$. Note that *tfu* is not the upper bound of the *mtfu*, and *mtfu* is the upper bound of each fuzzy itemset.

Definition 16: The start transaction period STP_{all} of all items is represented as

$$min_{TP} = \{STP_1, STP_2, \ldots, STP_m\}$$

The STP_j is the start transaction period of the j-th item, value m is the item number, and the symbol min_{TP} is assigned the latest time period of the parameters attached. For instance, in Table 1, STP_{all} is P_2.

Definition 17: Consider X be a temporal fuzzy utility itemset. LTP_{all} represents the time period from STP_{all} to the last time period of *TQD*. The temporal fuzzy utility upper bound ratio is represented as

$$tfuubr_X = \sum_{X \in Trans_y \cap Trans_y \in LTP_X} mtfu_{yX} \Big/ \sum_{Trans_y \in LTP_{all}} tfu_y$$

For instance, $tfuubr_{\{B.Low\}} = (mtfu_2 + mtfu_4) / (tfu_3 + tfu_4 + tfu_5 + tfu_6) = (21.44 + 45.42) / (1.98 + 53.34 + 47.36 + 26) = 66.86 / 128.69 = 51.95\%$.

Definition 18: Consider X and λ be temporal fuzzy utility itemset and a threshold value, respectively. X is a high temporal fuzzy utility upper bound itemset (*HTFUUBI*) if $tfuubr_X >= \lambda$ for any itemsets X. For instance, the temporal fuzzy utility 1-itemset {*B.Low*} in the above example is an $HTFUUBI_1$.

B. Proposed Mining Approach

The detailed steps for the proposed algorithm are as follow:

Inputs:

(1) temporal quantitative database (*TQD*)
(2) m items in *TQD*
(3) membership functions for all items
(4) t time periods
(5) minimum temporal fuzzy utility threshold λ

Step 1: Convert the appearance time of each transaction in *TQD* to a time period.

Step 2: Get the STP_{i_z} of each item i_z in *TQD*.

Step 3: For each item i_z in a $Trans_y$, change its value v_{yz} into a fuzzy set f_{yz} represented as $\left(\frac{f_{yz1}}{R_{z1}} + \frac{f_{yz2}}{R_{z2}} + \ldots + \frac{f_{yzh}}{R_{zh}}\right)$, according to the membership functions for the is the membership of v_{yz} of R_{zl}.

quantities of the items, where R_{zl} is the l-th linguistic term of i_z, h is the total number of regions for i_z, and f_{yzl} is the membership of v_{yz} of R_{zl}.

Step 4: For each $Trans_{jy}$ in each period p_j of TQD:

(a) Calculate the fu_{jyzl} of the l-th fuzzy region of item i_z in $Trans_{jy}$.

(b) Determine the mfu_{jyz} of i_z in $Trans_{jy}$.

(c) Find the tfu_{jy} value and the $mtfu_{jy}$ of $Trans_{jy}$.

Step 5: Initialize the $HTFUUBI_1$ table as empty, where each tuple has a fuzzy 1-itemset, the total *mtfu* value, and the occurrence frequency of the fuzzy 1-itemset.

Step 6: Derive the *tfuubr* of each temporal fuzzy utility 1-itemset. If $tfuubr \geq \lambda$, where λ is the given threshold value, the corresponding fuzzy 1-itemset is $HTFUUBI_1$.

Step 7: Insert the temporal fuzzy utility 1-itemset where $tfuubr \geq \lambda$ into the $HTFUUBI_1$ table. Include their total *mtfu* value and the occurrence frequency.

Step 8: Delete the temporal fuzzy utility 1-itemsets and those fuzzy utility values not included in the $HTFUUBI_1$ table.

Step 9: Sort the $HTFUUBI_1$ table from high to low frequency. This $HTFUUBI_1$ table can be used as the header table of the FP-Tree.

Step 10: Replace the values of the nodes of the FP-Tree with the *mtfu*, *TID* (transaction identification), fuzzy value and utility value from the transaction table.

Step 11: After completing step 10, the final FP-Tree is built. Then, using FP-Growth [8] to find the candidate *HTFUIs* and determine whether they are *HTFUIs* while mining the conditional FP-Tree.

Step 12: Output the *HTFUIs* to users.

IV. EXPERIMENTS

A synthetic dataset, T6I4N4KD200K, was used for compared performance efficiency between the proposed method and two-phased method [6] in terms of running time and consumed memory. The quantity value for each item in this dataset was assigned randomly in the range from [1, 10]. The profit values of items were also produced at random. Three membership functions in Figure 1 were used in the test.

The running time and consumed memory of the proposed method and the previous two-phased method for three different thresholds with three fuzzy terms for items are shown as follows. In Figures 2 and 3, it can be observed that the proposed method outperformed the two-phased method in execution time and consumed memory.

V. CONCLUSIONS

In this paper, we propose an extended tree structure according to the method in [6] to store more information in nodes and mine the results in one phase. The designed algorithm can find all high temporal fuzzy utility itemsets. The experimental results show that the proposed method can get good performance with less memory and with faster execution speed than the previous one.

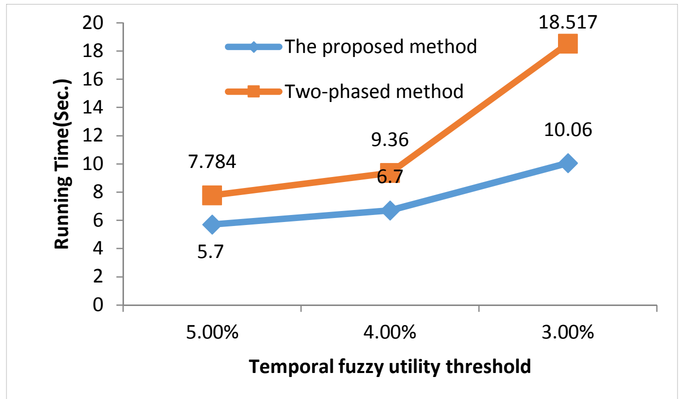

FIGURE 2. RUNNING TIME FOR TWO METHODS ON THREE DISTINCT THRESHOLDS.

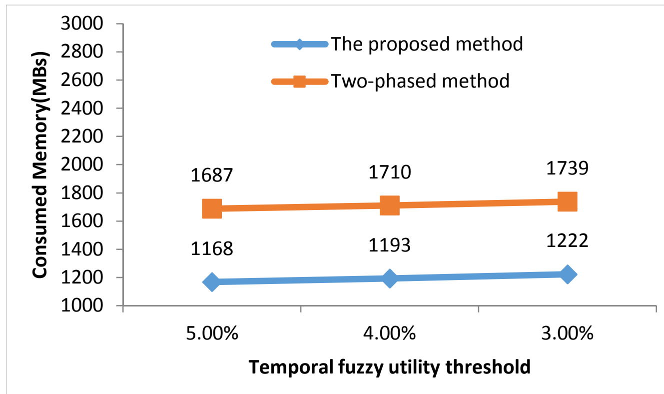

FIGURE 3. CONSUMED MEMORY FOR TWO METHODS ON THREE DISTINCT THRESHOLDS.

ACKNOWLEDGEMENT

This research was supported by the Ministry of Science and Technology in Taiwan under grants MOST 107-2221-E-390-016-MY2 and MOST 108-2218-E-390-001.

REFERENCES

[1] R. Agrawal and R. Srikant, “Fast algorithms for mining association rules,” *The International Conference on Very Large Data Bases*, Vol. 1215, pp. 487- 499, 1994.

[2] R. Agrawal, T. Imieliński, and A. Swami, “Mining association rules between sets of items in large database,” *The ACM SIGMOD International Conference on Management of Data*, Vol. 22, pp. 207-216, 1993.

[3] R. Agrawal, T. Imieliński, and A. Swami, “Database mining: a performance perspective,” *IEEE Transactions on Knowledge and Data Engineering*, Vol. 5, pp. 914-925, 1993.

[4] R. Chan, Q. Yang, and Y. D. Shen, “Mining high utility itemsets,” *The 3rd IEEE International Conference on Data Mining*, pp. 19-26, 2003.

[5] T. P. Hong, P. H. Chiang, W. M. Huang and J. C. W. Lin, “Fuzzy utility mining under minimum weight constraint of multiple item weights,” *The International Conference on Data Mining Workshops*, pp. 982-986, 2019.

[6] T. P. Hong, C. Y. Lin, W. M. Huang, S. M Li, S. L. Wang and J. C. W. Lin, “Mining temporal fuzzy utility itemsets by tree structure,” The IEEE International Conference on Big Data, 2019.

[7] B. Y. Wang and S. M. Zhang, “A mining algorithm for fuzzy weighted association rules,” *The Second International Conference on Machine Learning and Cybernetics*, pp. 2495-2499, 2003.

[8] J. Han, J. Pei, and Y. Yin, “Mining frequent patterns without candidate generation,” *The ACM SIGMOD International conference on management of data*, Vol. 29, pp. 1-12, 2000.

[9] J. C. W. Lin, W. Gan, P. Fournier-Viger, T. P. Hong, and H. C. Chao, “Mining weighted frequent itemsets without candidate generation in uncertain databases,” *International Journal of Information Technology & Decision Making*, vol. 16, no. 06, pp. 1549-1579, 2017.

[10] W. M. Huang, T. P. Hong, G. C. Lan, M. C. Chiang, and J. C. W. Lin, "Temporal-based Fuzzy Utility Mining," *IEEE Access*, Vol. 5, pp. 26639-26652, 2017.

[11] W. G. Huo, X. J. Feng, and Z. Y. Zhang, “An efficient approach for incremental mining fuzzy frequent itemsets with FP-Tree,” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 24, pp. 367-386, 2016.

[12] IBM Quest Data Mining Project, Quest synthetic data generation code. Available at (http://www.almaden.ibm.com/cs/quest/syndata. Html).

[13] G. C. Lan, T. P. Hong, and V. S. Tseng, “Discovery of high utility itemsets from on-shelf time periods of products,” *Expert Systems with Applications*, Vol. 38 , pp. 5851-5857, 2011.

[14] G. C. Lan, T. P. Hong, Y. H. Lin, and S. L. Wang, “Fuzzy utility mining with upper bound measure,” *Applied Soft Computing*, Vol. 30, pp. 767-777, 2015.

[15] C. W. Lin, T. P. Hong, and W. H. Lu, “Linguistic data mining with fuzzy FP-trees,” *Expert Systems with Applications*, Vol. 37, pp. 4560-4567, 2010.

[16] Y. Liu, W. Liao, and A. Choudhary, "A fast high utility itemsets mining algorithm," *The First International Workshop on Utility-Based Data Mining*, pp. 90-99, 2005.

[17] Y. Liu, W. Liao, and A. Choudhary, "A two-phase algorithm for fast discovery of high utility itemsets," *The Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Vol. 3518, pp. 689-695, 2005.

[18] Y. Qiu, Y. J. Lan, and Q. S. Xie, “An improved algorithm of mining from FP-tree,” *The International Conference on Machine Learning and Cybernetics*, Vol. 3, pp. 1665-1670, 2004.

[19] S. G. Totad, R. B. Geeta, and P. P. Reddy, “Batch incremental processing for FP-tree construction using FP-Growth algorithm,” *Knowledge and Information Systems*, Vol. 33, pp. 475-490, 2012.

[20] X. Wang, and L. Wang, “Incremental mining of high utility co-locations from spatial database,” *IEEE International Conference on Big Data and Smart Computing*, pp. 215-222, 2017.

[21] H. Yao, H. J. Hamilton, and C. J. Butz, “A foundational approach to mining itemset utilities from databases,” *The Third SIAM International Conference on Data Mining*, pp. 482-486, 2004.

[22] J. S. Yeh, Y. C. Li, and C. C. Chang, “Two-phase algorithm for a novel utility-frequent mining model,” *The Pacific Asia Conference on Knowledge Discovery and Data Mining*, Vol. 4819, pp. 433-444, 2007.

[23] L. A. Zadeh, “Fuzzy sets,” *Information and Control*, Vol. 8, pp. 338-353, 1965.

[24] C. H. Chen, B. Y. Chiang, T. P. Hong, D. C. Wang, J. C. W. Lin and M. Gankhuyag, “A fuzzy GGA-based approach to speed up evolutionary process for diverse group stock portfolio optimization,” *Journal of Intelligent & Fuzzy Systems*, Vol. 37, no. 6, pp. 7465-7479, 2019.

[25] O. Castillo, F. Kutlu and O. Atan, “Intuitionistic fuzzy control of twin rotor multiple input multiple output systems,” *Journal of Intelligent & Fuzzy Systems*, Vol. 38, no. 1, pp.821-833, 2020.

[26] Z. X. Wu, K. W. Huang, J. L. Chen and C. S. Yang, “A memetic fuzzy whale optimization algorithm for data clustering,” The 2019 IEEE Congress on Evolutionary Computation, pp. 1446-1452, 2019.

[27] G. C Lan, C. H. Chen, T. P. Hong and S. B. Lin, “A fuzzy approach for mining general temporal association rules in a publication database,” *The 11th International Conference on Hybrid Intelligent Systems*, pp. 611-615, 2011.

[28] C. H. Weng and Y. L. Chen, “Mining fuzzy association rules from uncertain data,” *Knowledge and Information Systems*, Vol. 23, no. 2, pp. 129-152, 2010.