

Fuzzy utility mining with upper-bound measure

Guo-Cheng Lan^a, Tzung-Pei Hong^{b,c,*}, Yi-Hsin Lin^b, Shyue-Liang Wang^d

^a Computational Intelligence Technology Center, Industrial Technology Research Institute, Hsinchu 310, Taiwan

^b Department of Computer Science and Information Engineering, National University of Kaohsiung, Kaohsiung 811, Taiwan

^c Department of Computer Science and Engineering, National Sun Yat-sen University, Kaohsiung 804, Taiwan

^d Department of Information Management, National University of Kaohsiung, Kaohsiung 811, Taiwan

ARTICLE INFO

Article history:

Received 3 February 2013

Received in revised form 27 January 2015

Accepted 27 January 2015

Available online 16 February 2015

Keywords:

Data mining

Fuzzy data mining

Fuzzy utility mining

High fuzzy utility itemset

Upper bound

ABSTRACT

Fuzzy utility mining has been an emerging research issue because of its simplicity and comprehensibility. Different from traditional fuzzy data mining, fuzzy utility mining considers not only quantities of items in transactions but also their profits for deriving high fuzzy utility itemsets. In this paper, we introduce a new fuzzy utility measure with the fuzzy minimum operator to evaluate the fuzzy utilities of itemsets. Besides, an effective fuzzy utility upper-bound model based on the proposed measure is designed to provide the downward-closure property in fuzzy sets, thus reducing the search space of finding high fuzzy utility itemsets. A two-phase fuzzy utility mining algorithm, named *TPFU*, is also proposed and described for solving the problem of fuzzy utility mining. At last, the experimental results on both synthetic and real datasets show that the proposed algorithm has good performance.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Data mining on knowledge discovery is a very important phase for extracting interesting rules and patterns from various data applications. Traditional association-rule mining was first proposed to find the relationship between items from a set of data [1,2], but transactions in retail databases usually included not only bought items but also the bought quantities of the items. Since traditional association-rule mining techniques were not sufficient to be used to handle such quantitative data, Srikant and Agrawal then proposed a new issue named quantitative association rule mining, which partitioned several attribute value ranges for each attribute to find useful quantitative rules. For example, assume a quantitative rule “{age: [20~29] → Car: [0,1]}”, which most customers with age ranging 20–29 usually do not buy cars or buy one car [21]. However, how to decide the suitable intervals for the domain values in each attribute is difficult, and the discovered rules are not easily to be comprehended by decision makers.

Afterward, Chan et al. thus proposed another new issue, namely utility mining, which considered not only quantities of items in transactions but also profits of items in a database, to find itemsets

with high utility values in databases [5]. Recently, utility mining has been an emerging issue in data mining due to its practical applications, such as medical data application, mobile data application, behavior data application, stream data application, etc. But, there is a big challenge in utility mining. That is, downward-closure property in association-rule mining cannot be applied to handle the problem of utility mining. To address this, Liu et al. subsequently proposed the two-phase utility mining algorithm (abbreviated as *TP*), which consisted of two main phases to find high utility itemsets in transaction databases [20]. In the first phase, an effective model, called transaction-weighted utilization (abbreviated as *TWU*), was proposed to construct a new downward-closure property [20]. The main principle was that the transaction utility (*tu*) of a transaction was regarded as the upper-bound of any subset in that transaction. The transaction-weighted utility (*twu*) of an itemset was then the summation of transaction utility values of the transactions including the itemset in the database. Next in the second phase, a predefined minimum utility threshold was used to find out itemsets with high utilities, and the itemsets could then be output as auxiliary information. According to definitions in utility mining, however, the output information for a high utility itemset only provided its utility and the items in it for decision makers [20]. Then, we could not acquire more other information from the results discovered by utility mining techniques, such as quantity interval of each item in an itemset [20]. As traditional association-rule mining, in addition, the utility itemset results might be not easily to be comprehended by users.

* Corresponding author at: Department of Computer Science and Information Engineering, National University of Kaohsiung, Kaohsiung 811, Taiwan. Tel.: +886 75919191.

E-mail addresses: rfoheiy@gmail.com (G.-C. Lan), tphong@nuk.edu.tw (T.-P. Hong), cccgo123@yahoo.com.tw (Y.-H. Lin).

As above mentioned, Wang et al. then proposed a new issue, namely fuzzy utility mining, which combined fuzzy set theory with utility mining to find high fuzzy utility itemsets (HFU) from quantitative databases [22]. Moreover, Wang et al.'s also defined a fuzzy utility function to evaluate fuzzy utility of an item in databases. For example, assume there is a quantitative transaction {1A, 3B}, in which the symbols and the numbers represent the bought items and their sold quantities, respectively. Also, assume the same membership function with three fuzzy regions: *Low*, *Middle*, and *High*, is assumed for the two distinct items, A and B. With the membership in Fig. 1, the quantities of the two items in the transaction can be converted into the two fuzzy sets $f_A = \{1/A.Low, 0/A.Middle, 0/A.High\}$ and $f_B = \{0.6/B.Low, 0.4/B.Middle, 0/B.High\}$. In addition, assume the profits of the two items, A and B, are 1 and 8. Take the fuzzy itemset {A.Low, B.Low} as an example. For the fuzzy term {B.Low} of {A.Low, B.Low}, since the centroid value of the fuzzy region "Low" in Fig. 1 is 1, and its profit and membership value are 8 and 0.6, the fuzzy utility of {B.Low} can be calculated as $0.6 \times (1 \times 8)$, which is 4.8. Similarly, the fuzzy utility of {A.Low} can be calculated as $1 \times (1 \times 1)$, which is 1. The fuzzy utility of the itemset {A.Low, B.Low} in the transaction is the summation of the two fuzzy terms in the transaction, which is 5.8 ($=1 + 4.8$). Finally, the fuzzy utility of an itemset is the summation of fuzzy utilities of the itemset in all transactions including the itemset in a database. As this example notes, however, the common membership degree value for the fuzzy terms in an itemset in fuzzy set theory is not considered in the fuzzy utility function, and the calculated fuzzy utility values for itemsets are also not easily to be explained.

Due to above reasons, this work thus presents a new fuzzy utility function, which considers not only quantities and profits of items but also the minimum operator principle of fuzzy set theory, to evaluate the actual fuzzy utility of an itemset in a set of transactions. However, since downward-closure property in fuzzy itemset mining cannot be kept in fuzzy utility mining with the proposed function, the effective fuzzy utility upper-bound model (abbreviated as *FUUB*) is proposed to avoid information losing. Based on the *FUUB* model, in addition, this work also presents the two-phase fuzzy utility mining approach (abbreviated as *TPFU*). At last, experimental results reveal the proposed *TPFU* approach can have good performance in terms of execution efficiency under various parameter settings.

The rest of parts in this paper is organized as follows. The related works are reviewed in Section 2. The problem to be solved and definitions are described in Section 3. The execution details of the proposed *TPFU* approach are explained in Section 4. An example in Section 4.2 is given to describe how to perform the mining process by using the proposed *TPFU*. The experimental results are then showed in Section 5. Finally, conclusions and future work are discussed in Section 6.

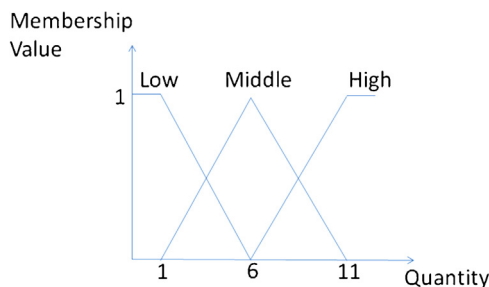


Fig. 1. The membership functions for all the four items in this example.

2. Review of related works

In this section, some related studies on association-rule mining, fuzzy data mining, and fuzzy utility mining are briefly reviewed.

2.1. Association-rule mining

In the field of data mining, discovery of association rules in large databases is one of the important issues due to the consideration of relationship of items [1,2]. For example, assume a product combination "{milks, breads}" has a high frequency in a database. It means the most customers usually buy milks and breads together in the retail. To deal with this problem, a famous algorithm, namely *Apriori*, was first proposed to find frequent itemsets in databases [1]. The *Apriori* algorithm consists of two main phases, finding frequent itemsets and finding association rules. In the first phase, all possible items in a database are regarded as candidate 1-itemsets, and the support of each item is scanned to count the support values of the candidates. Then, the frequent 1-itemsets with supports being larger than or equal to a pre-defined minimum support threshold could be found from the set of candidate 1-itemsets. Next, candidate 2-itemsets with two items are generated from the set of frequent 1-itemsets, and then the process of both the scanning data and counting supports is executed again for the candidate 2-itemsets. The whole mining task can be terminated until no candidate itemsets can be generated in the next pass. After the whole mining process, the final set of frequent itemsets in the database is then found.

2.2. Fuzzy data mining

As mentioned previously, a transaction in retail databases usually also includes quantities bought of items in that transaction in addition to items in the transaction. Then, traditional frequent itemset mining techniques cannot be sufficient to be used to handle such data with quantitative values [8,21]. To deal with such data, Srikant et al. thus proposed a new research issue, named quantitative association rule mining, which partitioned several attribute value ranges for each attribute to find useful quantitative association rules [21]. However, there exist some challenges in quantitative rule mining, such as that it is difficult for how to determine the suitable value ranges for the domain values of each attribute, and it is not easily to be comprehended by users for the rules discovered by the quantitative rule mining techniques.

The fuzzy set theory has widely been used various intelligent systems because of its simplicity and comprehensibility to human reasoning. To reach the goal, Kuok et al. first proposed a new research issue, fuzzy data mining, which applied the concept of fuzzy set theory to the data mining [16]. The main concept behind the issue is that quantitative values in transactions are converted into linguistic regions by fuzzy theory, and a minimum operator in fuzzy theory is applied to obtain the overlap value (minimum value) of membership regions in different items. Different from traditional association rules and quantitative rules, the interesting knowledge with simplicity and comprehensibility for fuzzy data mining could be found from the set of transactions with linguistic regions. To handle this problem, in addition, Hong et al. proposed an effective *Apriori*-based mining algorithm, which adopted a minimum operator in fuzzy theory to count the scalar cardinality value for an itemset in a transaction, to find interesting fuzzy association rules [10]. In addition, Hong et al. proposed an advanced mining approach, which considered the trade-off problem between number of rules and the cost of computation time. The main concept is that fuzzy term with the highest fuzzy count for the items could be kept in the set of fuzzy frequent 1-itemsets, and then a great deal of fuzzy rules could thus be avoided or the execution efficiency could

be improved [11]. Several other fuzzy data mining studies have also been published [4,6–8,12,14,18,19], but they were all *Apriori*-based techniques. Thus, the existing approaches had to spend a good deal of time cost generating a large number of candidates and counting their actual fuzzy counts in transactions. Based on the reasons, it is thus a quite critical issue to efficiently find fuzzy rules in quantitative databases.

2.3. Fuzzy utility mining

In association-rule mining, only binary itemsets are considered [1,2]. In real-world applications, however, products bought in transactions contain both profits and quantities. In particular, some high-profit products may occur with low frequency in a transaction database. For example, both jewel and diamond have high utility values but may not be frequent product combinations when compared to food and drink in a transaction database. Thus high-profit but low-frequency itemsets may not be found by the traditional association-rule mining approaches. To deal with this, Chan et al. proposed utility mining to discover high utility itemsets from a transaction database [5]. In this study, a high utility itemset considers not only the quantities of the items in transactions, but also their individual profits. Formally, local transaction utility and external utility are used to measure the utility of an item. The local transaction utility of an item is directly obtained from the information stored in a transaction database, like the quantity of the item sold in a transaction. The external utility of an item, like its profit, is given by users. External utility often reflects user preferences, and can be represented by a utility table or a utility function. By using a transaction dataset and a utility table together, the discovered itemset is able to better match a user's expectations than if found by considering only the transaction dataset itself.

Traditional association-rule mining keeps the downward-closure property [1,2], but utility mining does not, and thus the latter is much harder than the former. Liu et al. proposed a two-phase utility mining algorithm (*TP*) to discover high utility itemsets from a database by adopting the downward-closure property [20], and this approach was named as the transaction-weighted utilization (*TWU*) model. Based on principle of two-phase utility mining algorithm, several other studies about utility mining have also been published [3,13,17,23,24].

In real-world applications, however, how to get suitable promotion quantities of products in a product combination is a quite important. But, an itemset in utility mining only provides both of items in that itemset and its utility information for decision makers. To address this, Wang et al. proposed a new research issue, fuzzy utility mining, which combined fuzzy set theory with utility mining, to find high fuzzy utility itemsets (*HFU*) [22]. In particular, a new fuzzy utility function was defined to evaluate fuzzy utility of an item by the corresponding linguistic region value and degree value in its membership function. According to the fuzzy utility function in [22], however, it was observed that the minimum operator in traditional fuzzy data mining is not considered to evaluate the common degree values of fuzzy terms in a fuzzy itemset.

As mentioned above, this motivates our exploration of the issue, which designs a suitable fuzzy utility function and develops an efficient fuzzy utility mining approach.

3. Problem statement and definitions

To describe the proposed fuzzy utility mining problem clearly, assume a quantitative transaction database (*QDB*) with 10 transactions is given in Table 1, in which each transaction consists of three features – transaction identification (*TID*), items purchased, and their quantities in the transaction. There are four items in the

transactions, respectively denoted as *A* to *D*. The value attached to each item in the corresponding slot is the quantity sold in a transaction. Also, assume the profit value of each item is showed in Table 2, and the same membership functions with the three fuzzy regions, *Low*, *Middle* and *High*, are given for the four items for simplicity.

According to the above example, a set of terms related to the proposed fuzzy utility mining approach is then defined as follows.

Definition 1. An itemset *X* is a subset of items; That is, $X \subseteq I$. If $|X| = r$, the itemset *X* is called an *r*-itemset. Here let $I = \{i_1, i_2, i_3, \dots, i_r\}$ be a set of all distinct items which may appear in transactions. For example, the itemset $\{AB\}$ contains 2 items and is called a 2-itemset.

Definition 2. A quantitative transaction (*Trans*) is composed of a set of purchased items with their quantities, and a quantitative transaction database *QDB* is composed of a set of quantitative transactions. That is, $QDB = \{Trans_1, Trans_2, \dots, Trans_y, \dots, Trans_z\}$, where $Trans_y$ is the *y*th quantitative transaction in *QDB* and *z* is the number of transactions.

Definition 3. The fuzzy set f_{yz} of the quantitative value v_{yz} of the *z*th item i_z in a $Trans_y$ can be represented by the given membership functions for the item i_z as:

$$f_{yz} = \left(\frac{f_{yz1}}{R_{z1}} + \frac{f_{yz2}}{R_{z2}} + \dots + \frac{f_{yzl}}{R_{zl}} + \dots + \frac{f_{yzh}}{R_{zh}} \right)$$

where *h* is the number of regions for the item i_z , R_{zl} is the *l*th fuzzy region (linguistic term) of i_z , and f_{yzl} is the fuzzy membership value of the quantitative value v_{yz} of the *z*th item i_z in the *l*th fuzzy region R_{zl} .

For example in Table 1, the quantitative value (=3) of item *D* in $Trans_6$ can be converted to $f_{6,D} = (0.6/D.Low, 0.4/D.Middle, 0/D.High)$ by using the given membership function of item *D* with the three fuzzy regions shown in Fig. 1.

Definition 4. The external utility $s(i)$ of an item *i* in *I* is the corresponding utility value of the item in a utility table. Note that the external utility of an item can be obtained by users' expectation or a utility function to reflect the importance of the item, and thus the $s(i)$ value is a positive integer value or positive decimal value. For example, the external utility $s(D)$ of item *D* is 3 from Table 2.

Definition 5. The fuzzy utility fu_{yzl} of the *l*th fuzzy region of an item i_z in a quantitative transaction $Trans_y$ is the external utility

Table 1
The set of 10 quantitative transactions in this example.

<i>TID</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>
$Trans_1$	11	0	0	0
$Trans_2$	2	0	1	1
$Trans_3$	0	1	0	0
$Trans_4$	8	0	1	2
$Trans_5$	6	0	0	2
$Trans_6$	7	0	1	3
$Trans_7$	0	2	3	0
$Trans_8$	2	0	1	1
$Trans_9$	0	1	0	0
$Trans_{10}$	10	0	0	0

Table 2
The pre-defined profit values of the four items in this example.

Item	Profit
<i>A</i>	1
<i>B</i>	8
<i>C</i>	5
<i>D</i>	3

$s(i_z)$ of i_z multiplied by the quantitative value v_{yz} and the fuzzy membership value f_{yzl} of v_{yz} in the l th fuzzy region R_{yzl} . That is,

$$fu_{yzl} = f_{yzl} * v_{yz} * s(i_z).$$

For example from Tables 1 and 2, the quantity and the profit (external utility) of item D in $Trans_6$ are 3 and 3, respectively, and the fuzzy membership value of the quantity of item D in the first fuzzy region ($R_{D,Low}$) is 0.6 according to the given membership functions of the item D (as shown in Fig. 1). Thus, $fu_{6,D,Low} = 0.6 * 3 * 3 = 5.4$.

Note that Definition 3 defines the fuzzy membership values of fuzzy regions and the fuzzy utility used here is derived according to a crisp multiplication operation on the fuzzy membership values of fuzzy regions. These derived utility values are actually attached to fuzzy regions and thus the term of fuzzy utility is used to represent that the values are derived from fuzzy regions. Similar terms are adopted in the following definitions for simplicity.

Definition 6. The fuzzy utility fu_{yX} of a fuzzy itemset X in a transaction $Trans_y$ is the summation of the fuzzy utility values of all the fuzzy terms in X in $Trans_y$. That is,

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

where v_{yz} , $s(i_z)$, and f_{yzl} represent the quantity of the item i_z in $Trans_y$, the profit value of i_z , and the membership value of the l th fuzzy region of the item i_z in X , respectively. The term f_{yX} represents the membership value of X , and f_{yX} can be calculated by the minimum of all the membership values of all the fuzzy terms in X in $Trans_y$ by using the minimum operator.

For example in Table 1, the membership values of the two fuzzy terms $C.Low$ and $D.Low$ in the fuzzy itemset $\{C.Low, D.Low\}$ in $Trans_6$ are 1 and 0.6, respectively, according to the given membership function. By using the minimum operator, the membership value for the 2-itemset $\{C.Low, D.Low\}$ in $Trans_6$ is thus 0.6. The fuzzy utility of the fuzzy 2-itemset $\{C.Low, D.Low\}$ in $Trans_6$ can thus be calculated as $0.6 * ((1 * 5) + (3 * 3))$, which is 8.4.

Definition 7. The actual fuzzy utility afu_X of a fuzzy itemset X in a quantitative transaction database QDB is the summation of all fuzzy utility values of X in all the transactions including X . That is,

$$afu_X = \sum_y fu_{yX},$$

where fu_{yX} is the fuzzy utility of the fuzzy itemset X in the y th transaction $Trans_y$ in QDB .

For example in Table 1, $afu_{\{C.Low, D.Low\}} = fu_{2,\{C.Low, D.Low\}} + fu_{4,\{C.Low, D.Low\}} + fu_{6,\{C.Low, D.Low\}} + fu_{8,\{C.Low, D.Low\}} = 33.2$.

Definition 8. Let λ be a pre-defined minimum fuzzy utility threshold. A fuzzy itemset X is called a high fuzzy utility itemset (abbreviated as *HFU*) if $afu_X \geq \lambda$. For example in Table 1, if $\lambda = 25$ then the fuzzy 2-itemset $\{C.Low, D.Low\}$ is a high fuzzy utility 2-itemset because $afu_{\{C.Low, D.Low\}} = 33.2$, which is larger than λ .

Based on the above definitions, a fuzzy utility itemset considers not only the individual profits and quantities of items in quantitative transaction data but also the membership values of quantities of items. However, the fuzzy utility mining problem does not have the downward-closure property of traditional association-rule mining. For example, assume the threshold λ is set at 25. Since $afu_{\{D.Low\}} = 21 < \lambda$, $\{D.Low\}$ is a low utility fuzzy item in Table 1. However, its superset $\{C.Low, D.Low\}$ is a high fuzzy utility itemset because its $afu_{\{C.Low, D.Low\}}$ is 33.2, larger than λ . This example thus shows the downward-closure property cannot be hold for the problem. The proposed fuzzy utility mining is thus much harder than the traditional association-rule mining. To deal with this, we propose an effective upper-bound model, which is named fuzzy

utility upper-bound (abbreviated as *FUUB*), to achieve the goal. A set of related terms in the proposed *FUUB* model is stated below.

Definition 9. The maximal fuzzy utility mfu_{yz} of an item i_z in $Trans_y$ is defined as follows:

$$mfu_{yz} = \max\{fu_{yz1}, fu_{yz2}, \dots, fu_{yzl}\},$$

where fu_{yzl} is the fuzzy utility value of the l th fuzzy region R_{zl} of the item i_z in $Trans_y$. For example, according to Table 1 and Fig. 1, the membership values of the three regions (*Low*, *Middle*, and *High*) of item C in $Trans_7$ can be found as 0.6, 0.4 and 0, respectively, and the quantitative and the profit of item C in $Trans_7$ are 3 and 5, respectively. Then, $fu_{7,\{C.Low\}} = 0.6 * 3 * 5 = 9$, $fu_{7,\{C.Middle\}} = 0.4 * 3 * 5 = 6$, and $fu_{7,\{C.High\}} = 0 * 3 * 5 = 0$. Thus, $mfu_{7,\{C.Low\}} = 9$.

Definition 10. The maximal transaction fuzzy utility $mtfu_y$ of a quantitative transaction $Trans_y$ is the summation of the maximal fuzzy utility values of all the items in $Trans_y$. That is,

$$mtfu_y = \sum_{i_z \in Trans_y} mfu_{yz},$$

where mfu_{yz} is the maximal fuzzy utility of the z th item i_z in $Trans_y$. For example in $Trans_7$, $mtfu = mfu_{6,\{B.Low\}} + mfu_{6,\{C.Low\}} = 12.8 + 9 = 21.8$.

Definition 11. The fuzzy utility upper-bound $fuub_X$ of a fuzzy itemset X is the summation of the maximal transaction fuzzy utility values of all the transactions including X in QDB . That is,

$$fuub_X = \sum_{X \subseteq Trans_y, Trans_y \in QDB} mtfu_y.$$

For example in Table 1, $fuub_{\{C.Low\}} = mtfu_2 + mtfu_4 + mtfu_6 + mtfu_7 + mtfu_8 = 9.6 + 14.6 + 16 + 21.8 + 9.6 = 71.6$.

Definition 12. Let λ be a pre-defined minimum fuzzy utility threshold. A fuzzy itemset X is called a high fuzzy utility upper-bound itemset if $fuub_X \geq \lambda$. For example in Table 1, the fuzzy itemset $\{C.Low\}$ is a high fuzzy utility upper-bound 1-itemset.

3.1. Problem statement

The problem of fuzzy utility mining to be solved in this study is defined as follows. Assume a quantitative transaction database QDB contains a number of quantitative transactions, and each transaction is recorded with the purchased items and quantities. Besides, a utility table with the profits of the items is given. The problem to be solved here is to find all the fuzzy itemsets with their actual fuzzy utility values being larger than or equal to a predefined minimum fuzzy utility threshold λ . In this paper, a two-phase fuzzy utility mining algorithm (abbreviated as *TPFU*) is proposed to effectively and efficiently discover high fuzzy utility itemsets from a given quantitative transaction database QDB .

4. The proposed algorithm

In this section, we propose an effective upper-bound model and an efficient algorithm to find high fuzzy utility itemsets with the consideration of the fuzzy minimum operator. The details of the proposed two-phase fuzzy utility mining algorithm (abbreviated as *TPFU*) are then described as follows.

4.1. The proposed two-phase fuzzy utility mining algorithm, *TPFU*

This section introduces the proposed *TPFU* algorithm for efficiently finding high fuzzy utility itemsets in a quantitative

INPUT: A set of items, each with a profit value and a set of membership functions, a quantitative transaction database QDB , in which each transaction includes a subset of items with quantities, and the minimum fuzzy utility threshold λ .

OUTPUT: A final set of high fuzzy utility itemsets, $HFUs$.

Phase 1: Finding high fuzzy utility upper-bound itemsets

```

1  For each  $y$ -th transaction  $Trans_y$  in  $QDB$ {
2      Convert the quantitative value  $v_{yz}$  of each item  $i_z$  in  $Trans_y$  to a fuzzy set  $f_{yz}$ ;
3  }
4  For each  $y$ -th transaction  $Trans_y$  in  $QDB$ {
5      Find the fuzzy utility  $fu_{yz}$  of each fuzzy region  $R_z$  of each item  $i_z$  in  $Trans_y$ ;
6      Find the maximal fuzzy utility  $mfu_{yz}$  of each item  $i_z$  in  $Trans_y$ ;
7      Calculate the maximal transaction fuzzy utility  $mtfu_y$  of  $Trans_y$ ;
8  }
9  For each fuzzy term  $i_{zj}$  in  $QDB$ {
10     Calculate the fuzzy utility upper-bound  $fuub_{zj}$  of the fuzzy term  $i_{zj}$ ;
11 }
12 For each fuzzy term  $i_{zj}$  in  $QDB$ {
13     Find the set of high fuzzy utility upper-bound 1-itemsets,  $HFUUB_1$ ;
14 }
15 Set  $r = 1$ , where  $r$  represents the number of items in the itemsets to be processed;
16 Generate the candidate set  $C_{r+1}$  by the set of  $HFUUB_r$ ;
17 For each itemset  $X$  in the set  $C_{r+1}$ , find its fuzzy utility upper-bound  $fuub_X$  in  $QDB$ ;
18 Find the set of high fuzzy utility upper-bound  $(r+1)$ -itemsets  $HFUUB_{r+1}$  from the set  $C_{r+1}$ ;
19 If  $HFUUB_{r+1}$  is null, do STEP 21; otherwise, set  $r = r + 1$  and repeat STEPS 16 to 20.

```

Phase 2: Finding high fuzzy utility itemsets

```

20 For each itemset  $X$  in all  $HFUUB$  sets, find its actual fuzzy utility  $afu_X$  in  $QDB$ ;
21 Find all high fuzzy utility itemsets ( $HFUs$ ) in all  $HFUUB$  sets;
22 Output the final set of all high fuzzy utility itemsets,  $HFUs$ ;

```

Fig. 2. The execution procedure of the proposed $TPFU$ algorithm.

transaction database QDB . Fig. 2 shows the execution procedure of the proposed $TPFU$ algorithm.

The inputs of $TPFU$ include: (1) a quantitative transaction database, QDB , (2) a utility table for each item with a profit value $s(i)$, (3) a set of membership functions for each item and (4) a minimum fuzzy utility threshold λ . The output is the set of high fuzzy utility itemsets ($HFUs$) in QDB . The main phases of the $TPFU$ algorithm consist of two phases, (1) finding high fuzzy utility upper-bound itemsets and (2) finding high fuzzy utility itemsets. The key two phases of the $TPFU$ algorithm are executed and described as follows.

In the first phase, when the input data are fed, the quantitative value v_{yz} of each item i_z in each transaction $Trans_y$ is converted to a fuzzy set f_{yz} (Lines 01–03), and then the maximal transaction fuzzy utility ($mtfu_y$) of each transaction $Trans_y$ in QDB is first found by the maximal fuzzy utility mfu_{yz} of each item i_z in $Trans_y$ (Lines 04–08). Next, the fuzzy utility upper-bounds ($fuub$) of all distinct fuzzy terms in QDB are calculated (Lines 09–11), and the sets of high fuzzy utility upper-bound 1-itemsets ($HFUUB_1$) are found by their $fuub$ values (Lines 12–14). For now, the level-wise technique is adopted to generate candidate fuzzy $(r+1)$ -itemsets in the next pass by the current set of $HFUUB_r$ (Line 16), and the transactions in QDB are scanned to calculate and find the $fuub$ value of each

candidate $(r+1)$ -itemset (Line 17). After that, the set of high fuzzy utility upper-bound $(r+1)$ -itemsets ($HFUUB_{r+1}$) are found by their $fuub$ values (Line 18). The above process (Lines 16–18) is repeatedly done until no candidate $(r+1)$ -itemsets can be generated by the current set of $HFUUB_r$ (Line 19).

Next in the second phase, the data scan is executed once to find the actual fuzzy utility values of all itemsets in all $HFUUB$ sets (Line 20). Finally, all the high fuzzy utility itemsets ($HFUs$) can be found by their $fuub$ values, and the final set of all $HFUs$ are outputted as users' auxiliary information in making decisions (Lines 21–22).

4.2. An example of $TPFU$

In this section, an example is given to show how the proposed algorithm can find high fuzzy utility itemsets from a quantitative transaction database. Assume there are 10 transactions shown in Table 1 for mining, and there are four items in the transactions, respectively denoted as A to D . Also, assume the profit value of each item is shown in Table 2, and the same membership function with three fuzzy regions: *Low*, *Middle*, and *High*, is assumed for the four items to simplify and describe the problem of the fuzzy utility mining. Moreover, the minimum fuzzy utility threshold λ is set as

Table 3

The converted 10 quantitative transactions in this example.

<i>TID</i>	<i>A.L</i>	<i>A.M</i>	<i>A.H</i>	<i>B.L</i>	<i>B.M</i>	<i>B.H</i>	<i>C.L</i>	<i>C.M</i>	<i>C.H</i>	<i>D.L</i>	<i>D.M</i>	<i>D.H</i>
<i>Trans</i> ₁	0	0	1	0	0	0	0	0	0	0	0	0
<i>Trans</i> ₂	0.8	0.2	0	0	0	0	1	0	0	1	0	0
<i>Trans</i> ₃	0	0	0	1	0	0	0	0	0	0	0	0
<i>Trans</i> ₄	0	0.6	0.4	0	0	0	1	0	0	0.8	0.2	0
<i>Trans</i> ₅	0	1	0	0	0	0	0	0	0	0.8	0.2	0
<i>Trans</i> ₆	0	0.8	0.2	0	0	0	0	1	0	0.6	0.4	0
<i>Trans</i> ₇	0	0	0	0.8	0.2	0	0.6	0.4	0	0	0	0
<i>Trans</i> ₈	0.8	0.2	0	0	0	0	1	0	0	1	0	0
<i>Trans</i> ₉	0	0	0	1	0	0	0	0	0	0	0	0
<i>Trans</i> ₁₀	0	0.2	0.8	0	0	0	0	0	0	0	0	0

Table 4The fuzzy utilities of fuzzy regions of each item in *Trans*₂ in this example.

<i>TID</i>	<i>A.L</i>	<i>A.M</i>	<i>A.H</i>	<i>B.L</i>	<i>B.M</i>	<i>B.H</i>	<i>C.L</i>	<i>C.M</i>	<i>C.H</i>	<i>D.L</i>	<i>D.M</i>	<i>D.H</i>
<i>Trans</i> ₂	1.6	0.4	0	0	0	0	5	0	0	3	0	0

Table 5

The results for the maximal transaction fuzzy utility values of the 10 transactions in this example.

<i>TID</i>	<i>A.L</i>	<i>A.M</i>	<i>A.H</i>	<i>B.L</i>	<i>B.M</i>	<i>B.H</i>	<i>C.L</i>	<i>C.M</i>	<i>C.H</i>	<i>D.L</i>	<i>D.M</i>	<i>D.H</i>	<i>mtfu</i>
<i>Trans</i> ₁	0	0	1	0	0	0	0	0	0	0	0	0	11
<i>Trans</i> ₂	0.8	0.2	0	0	0	0	1	0	0	1	0	0	9.6
<i>Trans</i> ₃	0	0	0	1	0	0	0	0	0	0	0	0	8
<i>Trans</i> ₄	0	0.6	0.4	0	0	0	1	0	0	0.8	0.2	0	14.6
<i>Trans</i> ₅	0	1	0	0	0	0	0	0	0	0.8	0.2	0	10.8
<i>Trans</i> ₆	0	0.8	0.2	0	0	0	1	0	0	0.6	0.4	0	16
<i>Trans</i> ₇	0	0	0	0.8	0.2	0	0.6	0.4	0	0	0	0	21.8
<i>Trans</i> ₈	0.8	0.2	0	0	0	0	1	0	0	1	0	0	9.6
<i>Trans</i> ₉	0	0	0	1	0	0	0	0	0	0	0	0	8
<i>Trans</i> ₁₀	0	0.2	0.8	0	0	0	0	0	0	0	0	0	8

25. The proposed algorithm proceeds as follows to find high fuzzy utility itemsets from the quantitative transaction data in Table 1.

4.2.1. Phase 1: finding high fuzzy utility upper-bound itemsets

The quantity value of each item in each quantitative transaction is first converted into the corresponding a fuzzy set according to the membership functions of the items (Lines 01–03 in Fig. 2). Take the first quantitative transaction *Trans*₂ in Table 1 as an example. The transaction *Trans*₂ includes three items, A, C, and D, and their quantity values are 3, 1 and 1, respectively. Also, the same membership function for the three items, A, C, and D is given in Fig. 1. According to the membership function in Fig. 1, the quantity values of the three items in *Trans*₂ can be then converted into the three fuzzy sets, (0.8/*A.Low* + 0.2/*A.Middle* + 0/*A.High*), (1/*C.Low* + 0/*C.Middle* + 0/*C.High*), and (1/*D.Low* + 0/*D.Middle* + 0/*D.High*), respectively. The corresponding fuzzy set of the second transaction *Trans*₂ can be then consisted of (0.8/*A.Low* + 0.2/*A.Middle* + 0/*A.High*), (1/*C.Low* + 0/*C.Middle* + 0/*C.High*), and (1/*D.Low* + 0/*D.Middle* + 0/*D.High*). All other quantitative transactions in Table 1 can be similarly processed. After that, the results for the fuzzy sets of items of all transactions in Table 1 are then showed in Table 3. Note that “L”, “M”, and “H” in Table 3 are simplified as “Low”, “Middle”, and “High”.

For each transaction in Table 3, the fuzzy utility of each fuzzy region for each item in the transaction can be found (Lines 04–08 in Fig. 2). Take the item A in the second transaction *Trans*₂ in Table 3 as an example. The quantity of the item A in *Trans*₂ is 2, and its profit is 1. In addition, the membership values of the three fuzzy regions (*A.Low*, *A.Middle*, and *A.High*) of the item A, are 0.8, 0.2, and 0, respectively. The fuzzy utility values of the three regions of A in *Trans*₂ can be then calculated as 0.8*2*1 (=1.6), 0.2*2*1 (=0.4), and 0*2*1 (=0). The same process can be done for all the other items in the second transaction *Trans*₂. The results for the corresponding

fuzzy utility values of fuzzy regions of all items in *Trans*₂ are showed in Table 4.

Next, the maximal transaction fuzzy utility of the transaction can be found. Continuing the example of the second transaction *Trans*₂ in Table 4, the maximal fuzzy utility values in fuzzy utility of regions of the three items A, C, and D are found as 1.6, 5, and 3. The maximal transaction fuzzy utility of *Trans*₂ can be calculated as 1.6+5+3, which is 9.6. All the other transactions in Table 3 can be similarly processed. The results for maximal transaction fuzzy utility (*mtfu*) values of all transactions in Table 3 are shown in Table 5.

For each possible fuzzy terms in Table 5, the fuzzy utility upper-bound of the fuzzy item are found (Lines 09–11 in Fig. 2). Take the fuzzy term *A.Low* in Table 5 as an example. The fuzzy term, *A.Low*, appears in the two transactions, *Trans*₂ and *Trans*₈, and the maximal transaction fuzzy utility values of the two transactions are all 9.6. The fuzzy utility upper-bound of *A.Low* can then be calculated as 9.6+9.6, which is 19.2, and the *A.Low* and its upper-bound value are put in the set of candidate 1-itemsets, as shown in Table 6. All the other fuzzy items in Table 5 can be processed in the same way. The results for the fuzzy utility upper-bound (*fuub*) of all fuzzy terms in the set of *C*₁ are shown in Table 6.

Table 6All the fuzzy utility 1-itemsets in the set of *C*₁ in the example.

1-Itemset	<i>fuub</i>	1-Itemset	<i>fuub</i>
{ <i>A.Low</i> }	19.2	{ <i>C.Low</i> }	71.6
{ <i>A.Middle</i> }	68.6	{ <i>C.Middle</i> }	21.8
{ <i>A.High</i> }	49.6	{ <i>C.High</i> }	0
{ <i>B.Low</i> }	37.8	{ <i>D.Low</i> }	60.6
{ <i>B.Middle</i> }	21.8	{ <i>D.Middle</i> }	41.4
{ <i>B.High</i> }	0	{ <i>D.High</i> }	0

Table 7

The set of high fuzzy utility upper-bound 1-itemsets in this example.

1-Itemset	<i>fuub</i>
{A.Middle}	68.6
{A.High}	49.6
{B.Low}	37.8
{C.Low}	71.6
{D.Low}	60.6
{D.Middle}	41.4

The high fuzzy utility upper-bound 1-itemsets ($HFUUB_1$) are found from the set of candidate 1-itemsets (C_1) (Lines 12–14 in Fig. 2). Take the fuzzy 1-itemset {B.Low} in Table 6 as an example. The fuzzy utility upper-bound values of {B.Low} in Table 6 are found as 37.8, and the minimum fuzzy utility threshold λ is set at 25. For the fuzzy itemset {B.Low}, since its fuzzy utility upper-bound is larger than or equal to the minimum fuzzy utility threshold, it is put in the set of high fuzzy utility upper-bound 1-itemsets ($HFUUB_1$). The same process can be executed for all the other 1-itemsets in Table 6. The results for the sets of all high fuzzy utility upper-bound 1-itemsets ($HFUUB_1$) are showed in Table 7.

The variable r is currently set at 1, where r is used to represent the number of items in the current candidate fuzzy itemsets to be processed (Line 15 in Fig. 2).

In this example, thirteen candidate fuzzy utility 2-itemsets can be generated from the set $HFUUB_1$ as follows: {A.Middle, B.Low}, {A.Middle, C.Low}, {A.Middle, D.Low}, {A.Middle, D.Middle}, {A.High, B.Low}, {A.High, C.Low}, {A.High, D.Low}, {A.High, D.Middle}, {B.Low, C.Low}, {B.Low, D.Low}, {B.Low, D.Middle}, {C.Low, D.Low} and {C.Low, D.Middle} (Line 16 in Fig. 2). Note that there is a constraint for each valid fuzzy utility itemset. That is, it has to be different for all items without fuzzy regions in a fuzzy itemset. For example, the fuzzy utility 2-itemset {A.Middle, A.High} is not a valid one, but {A.Middle, B.Low} is.

The fuzzy utility upper-bound (*fuub*) values of the candidate itemsets to be processed can be found in the set of transactions in Table 5 (Line 17 in Fig. 2). Take the candidate fuzzy utility 2-itemset {A.Middle, C.Low} as an example. The candidate itemset appears in the four transactions, $Trans_2$, $Trans_4$, $Trans_6$, and $Trans_8$, and the maximal transaction fuzzy utility values of the four transactions are 9.6, 14.6, 16, and 9.6, respectively. The fuzzy utility upper-bound of the candidate 2-itemset can then be calculated as $9.6 + 14.6 + 16 + 9.6$, which is 49.8. The results for the fuzzy utility upper-bound values of all candidate 2-itemsets are shown in Table 8.

The high fuzzy utility upper-bound 2-itemsets ($HFUUB_2$) are found from the set of candidate 2-itemsets (C_2) (Line 18 in Fig. 2). The process is the same as that mentioned in STEP 4. In this example, there are eight high fuzzy utility upper-bound 2-itemsets ($HFUUB_2$) in the set of C_2 , which their fuzzy utility upper-bounds satisfy the minimum fuzzy utility threshold ($=25$), as shown in Table 9.

STEP 10: In this example, since set $HFUUB_2$ is not null, r is incremented to 2 and STEPs 6–10 are repeated (Line 19 in Fig. 2). The whole mining task can be terminated until no candidate itemsets can be generated in the next pass.

Table 8

The final results for the candidate fuzzy utility upper-bound 2-itemsets.

2-Itemset	<i>fuub</i>	2-Itemset	<i>fuub</i>
{A.Middle, B.Low}	0	{A.High, D.Middle}	30.6
{A.Middle, C.Low}	49.8	{B.Low, C.Low}	21.8
{A.Middle, D.Low}	60.6	{B.Low, D.Low}	0
{A.Middle, D.Middle}	41.4	{B.Low, D.Middle}	0
{A.High, B.Low}	0	{C.Low, D.Low}	49.8
{A.High, C.Low}	30.6	{C.Low, D.Middle}	30.6
{A.High, D.Low}	30.6		

Table 9

The set of high fuzzy utility upper-bound 2-itemsets in this example.

2-Itemset	<i>fuub</i>	2-Itemset	<i>fuub</i>
{A.Middle, C.Low}	49.8	{A.High, D.Low}	30.6
{A.Middle, D.Low}	60.6	{A.High, D.Middle}	30.6
{A.Middle, D.Middle}	41.4	{C.Low, D.Low}	49.8
{A.High, C.Low}	30.6	{C.Low, D.Middle}	30.6

Table 10

The actual fuzzy utilities of each high fuzzy utility upper-bound itemset.

Itemset	<i>afu</i>	Itemset	<i>afu</i>
{B.Low}	28.8	{A.High, D.Low}	8.8
{C.Low}	29	{A.High, D.Middle}	6
{A.Middle, B.Low}	0	{B.Low, C.Low}	18.6
{A.Middle, C.Low}	20.2	{B.Low, D.Low}	0
{A.Middle, D.Low}	29.6	{B.Low, D.Middle}	0
{A.Middle, D.Middle}	11.6	{C.Low, D.Low}	33.2
{A.High, B.Low}	0	{C.Low, D.Middle}	7.8
{A.High, C.Low}	7.6	{A.Middle, C.Low, D.Low}	17.8

4.2.2. Phase 2: finding high fuzzy utility itemsets

The actual fuzzy utility (*afu*) of each itemset in the set of all high fuzzy utility upper-bound itemsets found in the first phase can be found (Line 20 in Fig. 2). Take the candidate fuzzy utility 2-itemset {A.Middle, C.Low} as an example. The candidate itemset appears in the four transactions, $Trans_2$, $Trans_4$, $Trans_6$, and $Trans_8$, the membership values of the fuzzy item A.Middle in the four transactions are 0.2, 0.6, 0.8, and 0.2, and the values of C.Low are all 1. The actual fuzzy utility of {A.Middle, C.Low}, the minimum membership value of all items in the itemset in each transaction has to be first found. Continuing the example, take the second transaction $Trans_2$ in Table 5 as an example. Since the membership values of the two fuzzy items in $Trans_2$ are 0.2 and 1, respectively, the minimum value of the two membership values is 0.2. In addition, the quantities of the two items, A and C, in the transaction in Table 1 are 2 and 1, and the profits of the two items A and C in Table 2 are found as 1 and 5. Then, the fuzzy utility of the itemset {A.Middle, C.Low} in $Trans_2$ can be calculated as $0.2 \times (2 \times 1 + 1 \times 5)$, which is 1.4. The same process can be done for the other three transactions with the candidate 2-itemset {A.Middle, C.Low}, $Trans_4$, $Trans_6$, and $Trans_8$. Then, the result values of the three transactions are 7.8, 9.6, and 1.4. Thus, the actual fuzzy utility of the itemset {A.Middle, C.Low} can be calculated as $1.4 + 7.8 + 9.6 + 1.4$, which is 20.2. All the other twelve candidate 2-itemsets in the set of C_2 can be processed in the same fashion. The results for the actual fuzzy utility values of all candidate 2-itemsets are showed in Table 10.

All high fuzzy utility itemsets are found in Table 10 (Line 21 in Fig. 2). Take the fuzzy itemset {B.Low} in Table 10 as an example. The actual fuzzy utility of {B.Low} in Table 10 is 28.8, and the minimum fuzzy utility threshold λ is set at 25. Thus, {B.Low} can be put in the set of high fuzzy utility itemsets (HFU) because its actual fuzzy utility is larger than or equal to the minimum fuzzy utility threshold. The same process can be done for all the other itemsets in Table 10. After the step, the results for all high fuzzy utility itemsets (HFU) are showed in Table 11.

Table 11

The final set of all high fuzzy utility itemsets in this example.

Itemset	<i>afu</i>
{B.Low}	28.8
{C.Low}	29
{A.Middle, D.Low}	29.6
{C.Low, D.Low}	33.2

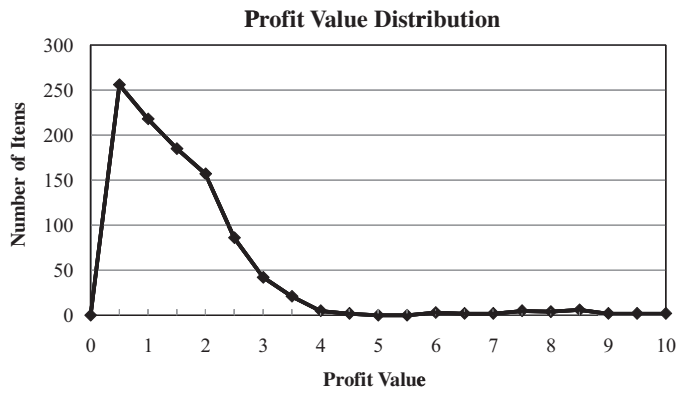


Fig. 3. The profit-value distribution in the experimental datasets.

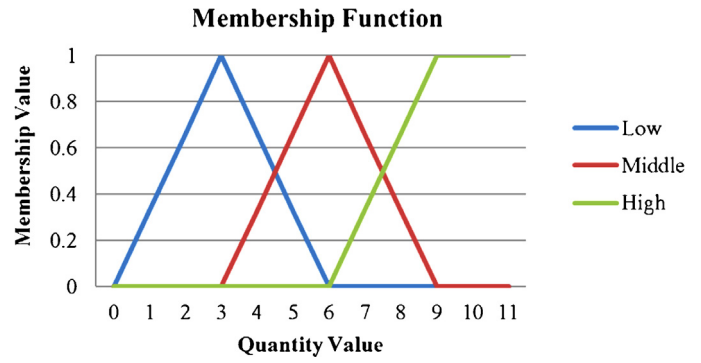


Fig. 4. The membership function used in the experiments.

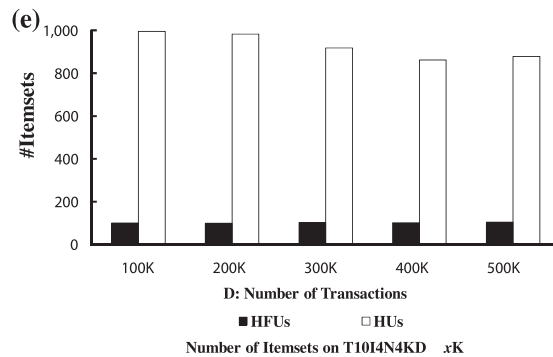
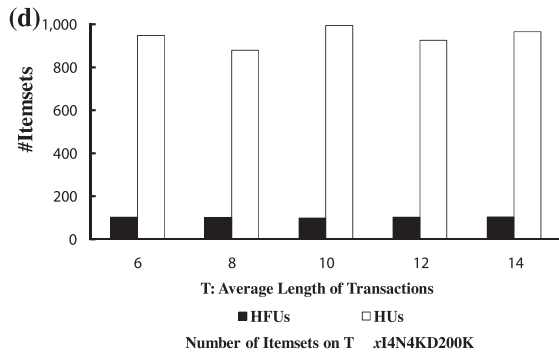
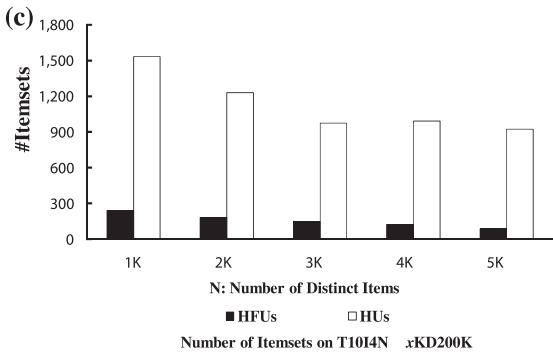
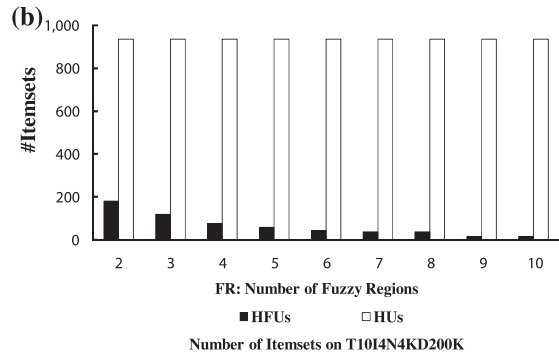
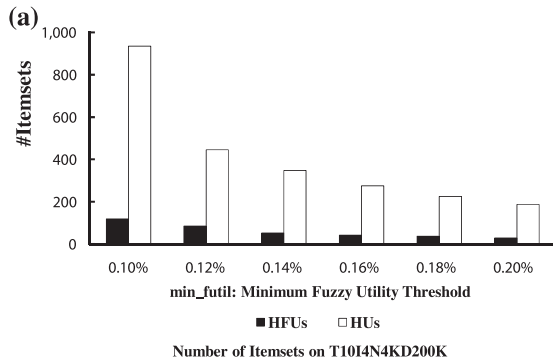


Fig. 5. Numbers of HFUs and HUs under various \min_futil , FR , N , T , and D .

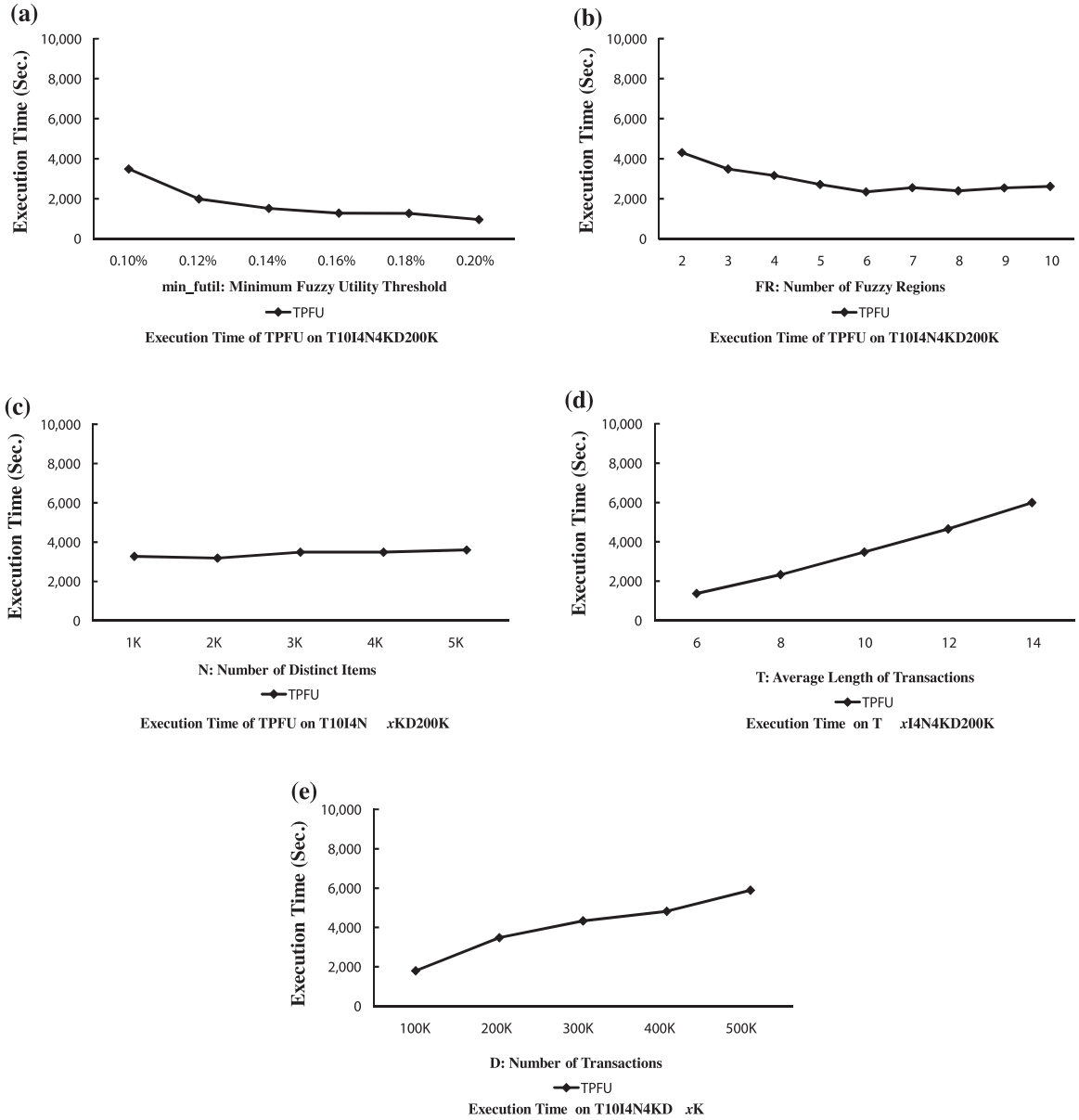


Fig. 6. Execution time of TPFU along with various min_futil , FR , N , T , and D .

In this example, the four high fuzzy utility itemsets in Table 11, $\{B.Low\}$, $\{C.Low\}$, $\{A.Middle, D.Low\}$, and $\{C.Low, D.Low\}$, are output as the decision makers' auxiliary information (Line 22 in Fig. 2).

As this example shows, the proposed upper-bound model can be used to effectively reduce the search space of find high fuzzy utility itemsets, and then all high fuzzy utility itemsets in a quantitative database can be found by using the proposed TPFU algorithm.

5. Experimental evaluation

A series of experiments were conducted to compare the performance of the proposed two-phase fuzzy utility algorithm (named TPFU) for various parameter values. The algorithm was implemented in J2SDK 1.6.0 and executed on a PC with a 3.30 GHz CPU and 1 GB memory.

5.1. Experimental datasets

To show the practical performance, a real public dataset (called BMS-POS) [9] was used in the experiments. This dataset was used in the KDDCUP 2000 competition, and it collected several years of point-of-sale data from a large electronics retailer, making it very suitable as a test dataset in the experiments. The details of this dataset were follows. Each transaction consisted of all the product categories purchased by a customer at one time. There were 515,597 transactions in this dataset and the total number of different items was 1657. In addition, the maximal length of a transaction was 164 and the average length of the transactions was 6.5.

On the other hand, the public IBM data generator was also used in our experiments to produce the data sets [15]. Since our purpose was to find out high fuzzy utility itemsets, we thus developed a simulation model, which was similar to that used in Liu et al.'s study [20], to generate the quantities of the items in the transactions. Each quantity ranged among 1–10 according to the described

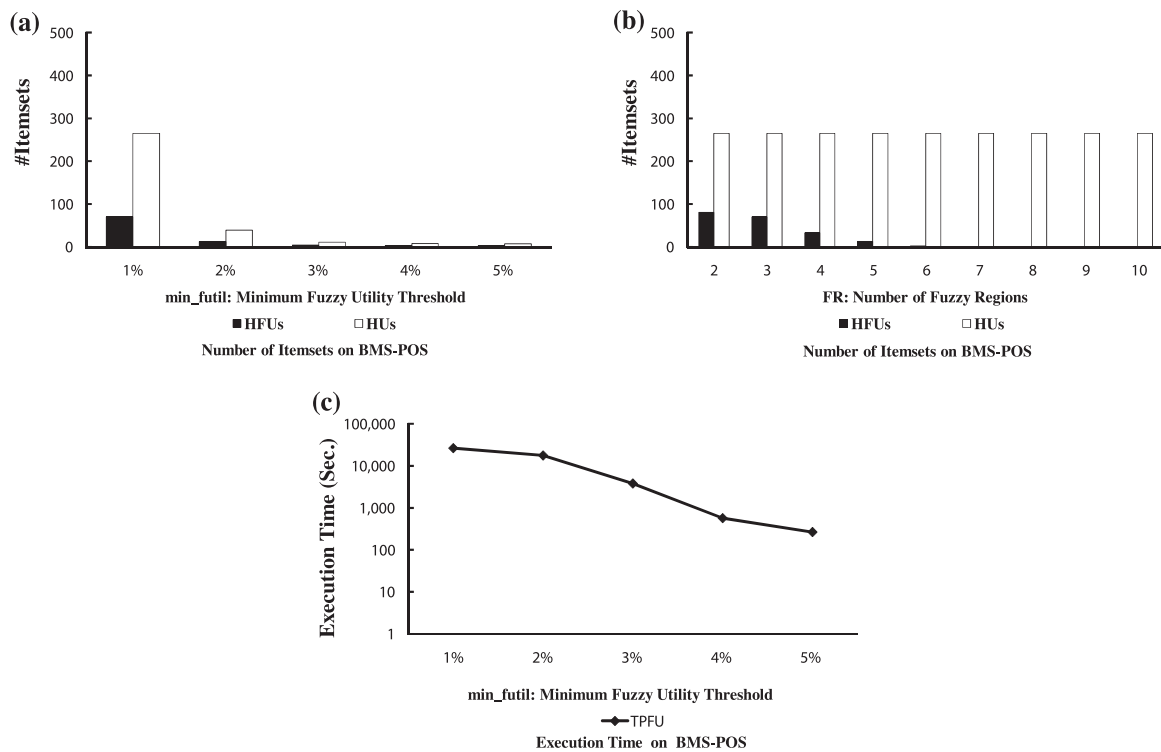


Fig. 7. Performance of TPFU along on the BMS-POS dataset.

way in [20]. Moreover, for each dataset generated, a corresponding utility table was also produced, in which a profit value using a log normal distribution in the range from 0.01 to 10.00 was randomly assigned to an item according to Liu et al.'s study [20]. The default values of parameters used in the experiments are listed in Table 12, and Fig. 3 showed the profit-value distribution of all the items generated by the simulation model in the utility table. Note that users can set a percentage value of the minimum fuzzy utility, and then the total utility of a database multiplied by the user-specified percentage value is the corresponding actual minimum fuzzy utility value (min_futil).

In the experiments, we assumed that all items had the same membership function, in which there are three fuzzy regions, *Low*, *Middle*, and *High*, as shown in Fig. 4.

5.2. Difference evaluation of HFUs and HUs

Experiments were first conducted on the synthetic datasets to evaluate the difference in the number of high fuzzy utility itemsets (HFUs) and high utility itemsets (HUs) under various parameters. Fig. 5(a)–(e) showed a comparison of the numbers of HFUs and HUs for the synthetic datasets for various min_futil , FR , T , N , and D , respectively. Note that FR represents the number of the fuzzy regions in a membership function.

It could clearly be observed from the figures that the number of HFUs is less than that of HUs in various parameters. The main reason is that our proposed fuzzy utility function is extended from

both the original utility function and minimum operator concept. That is, according to Definitions 6 and 7, the minimum value of fuzzy regions of all fuzzy terms in a fuzzy utility itemset in a transaction is multiplied by the original utility value of the itemset without consideration of minimum operator concept. Thus, the fuzzy utility of an itemset is less than the original utility value of that itemset. However, although the fuzzy utility of an itemset is less than the original utility value of that itemset, more useful quantitative information of fuzzy terms in that fuzzy utility itemset could be provided users because of its simplicity and comprehensibility when compared with itemsets in the traditional utility mining without consideration of the fuzzy set theory. Accordingly, the effectiveness of the mined high fuzzy utility itemsets can be showed in the figures.

5.3. Efficiency evaluation

Experiments were then on the synthetic datasets to evaluate the execution efficiency of the TPFU algorithms under parameter settings. Fig. 6(a)–(e) showed the execution efficiency of the proposed TPFU for the synthetic datasets under various varied min_futil , FR , N , T , and D , respectively.

As shown the figures, it could be observed that the execution efficiency of the proposed TPFU could still keep good efficiency to handle the problem of fuzzy utility mining with the help of the proposed upper-bound model. The reason for this is that based on the model, the proposed TPFU could effective reduce the search space of finding fuzzy utility itemsets in comparison with the full search space of itemsets. Hence, the proposed TPFU could perform the fuzzy utility mining task efficiently under various parameter settings.

5.4. Evaluation on real dataset BMS-POS

In the experiment, the real dataset BMS-POS [9] was then used to evaluate the performance of the proposed TPFU algorithm. Fig. 7 showed the number of two kinds of itemsets and execution

Table 12
Parameters used in the experiments.

Parameter	Description	Default
T	The average length of items per transaction	10
I	The average length of maximal potentially frequent itemsets	4
N	The total number of items	4000
D	The total number of transactions	200,000

efficiency of the *TPFU* for different thresholds, varying from 1% to 5%, respectively. Note that the *min.futil* value in Fig. 7(b) is set to the value of 1%, and the membership function with three fuzzy regions (as shown in Fig. 4) is set assumed in the two figures.

As shown in the figures, the number of the mined *HFUs* is less than that of the traditional *HUs*, and the proposed *TPFU* has stable efficiency in mining. The reason was the same as that mentioned in Sections 5.2 and 5.3. Accordingly, the *TPFU* algorithm has good performance on the real *BMS-POS* dataset in terms of effectiveness of *HFUs* and execution efficiency.

6. Conclusions

This work presents a new fuzzy utility function and an efficient two-phase fuzzy utility mining algorithm, named *TPFU*, to find high fuzzy utility itemsets in quantitative transaction databases. Different from the existing fuzzy utility algorithms, in particular, the defined fuzzy utility function in the *TPFU* considers not only the quantities and profits of items in transactions but also the common degrees of fuzzy terms in an itemset by using the concept of fuzzy minimum operator. In addition, an effective upper-bound model is developed to form the downward-closure property in fuzzy utility mining problem, and then all high fuzzy utility itemsets can be found in databases by using the proposed *TPFU*. The experimental results reveal the proposed *TPFU* algorithm based on the proposed upper-bound model can keep good performance in mining under different parameter settings when working with synthetic datasets and a public real dataset, *BMS-POS*.

In the future, we aim to apply the proposed approach to other practical applications, such as data streams. In addition, how to further improve the execution efficiency for mining high fuzzy utility itemsets is another interesting topic, such as the development of tree-based techniques. Besides, we will also attempt to address the maintenance problem of high fuzzy utility mining when transactions are inserted, deleted or modified.

References

- [1] R. Agrawal, R. Srikant, Fast algorithm for mining association rules, in: Proceedings of the International Conference on Very Large Data Bases, 1994, pp. 487–499.
- [2] R. Agrawal, T. Imielinski, A. Swami, Mining association rules between sets of items in large database, in: Proceedings of the ACM SIGMOD International Conference on Management of Data, 1993, pp. 207–216.
- [3] C.F. Ahmed, S.K. Tanbeer, B.S. Jeong, Y.K. Lee, Efficient tree structures for high utility pattern mining in incremental databases, *IEEE Trans. Knowl. Data Eng.* 21 (3) (2009) 1708–1721.
- [4] J. Alcalá-Fdez, R. Alcalá, M.J. Gacto, F. Herrera, Learning the membership function contexts for mining fuzzy association rules by using genetic algorithms, *Fuzzy Sets Syst.* 160 (7) (2009) 905–921.
- [5] R. Chan, Q. Yang, Y.D. Shen, Mining high utility itemsets, in: Proceedings of the 3rd IEEE International Conference on Data Mining, 2003, pp. 19–26.
- [6] C.H. Chen, T.P. Hong, V.S. Tseng, An improved approach to find membership functions and multiple minimum supports in fuzzy data mining, *Expert Syst. Appl.* 36 (6) (2009) 10011–10016.
- [7] C.H. Chen, T.P. Hong, V.S. Tseng, Fuzzy data mining for time-series data, *Appl. Soft Comput.* 12 (1) (2012) 536–542.
- [8] K.C.C. Chen, W.H. Au, An effective algorithm for mining interesting quantitative association rules, in: Proceedings of the 1997 ACM Symposium on Applied Computing, 1997, pp. 88–90.
- [9] Frequent Itemsets Mining Dataset Repository. Available at <http://fimi.cs.helsinki.fi/data/>
- [10] T.P. Hong, C.S. Kuo, S.C. Chi, A fuzzy data mining algorithm for quantitative values, in: Proceedings of the 3rd International Conference on Knowledge-Based Intelligent Information Engineering System, 1999, pp. 480–483.
- [11] T.P. Hong, C.S. Kuo, S.C. Chi, Trade-off between computation time and number of rules for fuzzy mining from quantitative data, *Int. J. Uncertain. Fuzz. Knowl.-Based Syst.* 9 (5) (2001) 587–604.
- [12] T.P. Hong, K.Y. Lin, S.L. Wang, Fuzzy data mining for interesting generalized association rules, *Fuzzy Sets Syst.* 138 (2) (2003) 255–269.
- [13] J. Hu, A. Mojsilovic, High-utility pattern mining: a method for discovery of high-utility item sets, *Pattern Recognit.* 40 (11) (2007) 3317–3324.
- [14] Tony C.K. Huang, Mining the change of customer behavior in fuzzy time-interval sequential patterns, *Appl. Soft Comput.* 12 (3) (2012) 1068–1086.
- [15] IBM Quest Data Mining Projection, Quest synthetic data generation code, 1996, Available at <http://www.almaden.ibm.com/cs/quest/syndata.htm>
- [16] C.M. Kouk, A. Fu, M.H. Wong, Mining fuzzy association rules in database, *ACM SIGMOD Rec.* 27 (1) (1998) 41–46.
- [17] G.C. Lan, T.P. Hong, Vincent S. Tseng, Discovery of high utility itemsets from on-shelf time periods of products, *Expert Syst. Appl.* 38 (5) (2011) 5851–5857.
- [18] Y.C. Lee, T.P. Hong, T.C. Wang, Multi-level fuzzy mining with multiple minimum supports, *Expert Syst. Appl.* 34 (1) (2008) 459–468.
- [19] J.H. Liu, Mining sequential patterns with fuzzy taxonomy structures, *J. Chienkuo Technol. Univ.* 24 (2) (2005) 193–210.
- [20] Y. Liu, W.K. Liao, A. Choudhary, A fast high utility itemsets mining algorithm, in: Proceedings of the Utility-Based Data Mining Workshop, 2005, pp. 90–99.
- [21] R. Srikant, R. Agrawal, Mining quantitative association rules in large relational tables, in: Proceedings of the ACM SIGMOD Conference on Management of Data, 1996, pp. 1–12.
- [22] C.M. Wang, S.H. Chen, Y.F. Huang, A fuzzy approach for mining high utility quantitative itemsets, in: Proceedings of the IEEE International Conference on Fuzzy Systems, 2009, pp. 20–24.
- [23] J.S. Yeh, P.C. Hsu, HHUIF and MSICF: novel algorithms for privacy preserving utility mining, *Expert Syst. Appl.* 37 (7) (2010) 4779–47786.
- [24] J.S. Yeh, C.Y. Chang, Y.T. Wang, Efficient algorithms for incremental utility mining, in: Proceedings of Conference on Ubiquitous Information Management and Communication, 2008, pp. 229–234.