Soft Comput (2014) 18:2413–2424
DOI 10.1007/s00500-013-1214-4

METHODOLOGIES AND APPLICATION

Actionable high-coherent-utility fuzzy itemset mining

Chun-Hao Chen · Ai-Fang Li · Yeong-Chyi Lee

Published online: 9 January 2014

Abstract Many fuzzy data mining approaches have been proposed for finding fuzzy association rules with the predefined minimum support from quantitative transaction databases. Since each item has its own utility, utility itemset mining has become increasingly important. However, common problems with existing approaches are that an appropriate minimum support is difficult to determine and that the derived rules usually expose common-sense knowledge, which may not be interesting from a business point of view. This study thus proposes an algorithm for mining high-coherent-utility fuzzy itemsets to overcome problems with the properties of propositional logic. Quantitative transactions are first transformed into fuzzy sets. Then, the utility of each fuzzy itemset is calculated according to the given external utility table. If the value is larger than or equal to the minimum utility ratio, the itemset is considered as a high-utility fuzzy itemset. Finally, contingency tables are calculated and used for checking whether a high-utility fuzzy itemset satisfies four criteria. If so, it is a high-coherent-utility fuzzy itemset. Experiments on the *foodmart* and simulated datasets are made to show that the derived itemsets by the proposed algorithm not only can reach better profit than selling them separately, but also can provide fewer but more useful utility itemsets for decision-makers.

Keywords Fuzzy set · Utility fuzzy itemsets · High-coherent-utility fuzzy itemset · Membership function · Domain-driven data mining

Communicated by Z. Zhu.

This is a modified and expanded version of the paper "A high coherent utility fuzzy itemsets mining algorithm," *The 2012 International Conference on Information Security and Intelligent Control*, pp. 114–117, 2012.

C.-H. Chen (✉) · A.-F. Li
Department of Computer Science and Information Engineering, Tamkang University, Tamsui, New Taipei, Taiwan, ROC
e-mail: chchen@mail.tku.edu.tw

A.-F. Li
e-mail: 697410172@s97.tku.edu.tw

Y.-C. Lee
Department of Information Management, Cheng Shiu University, Kaohsiung, Taiwan, ROC
e-mail: yeongchyi@csu.edu.tw

1 Introduction

Data mining is most commonly used to derive useful information and extract useful patterns from large datasets or databases. One commonly used technique is association rule mining, which is an expression $X \rightarrow Y$, where X and Y are sets of items (Agrawal and Srikant 1994). The expression means that in the set of transactions, if all the items in X exist in a transaction, then all items in Y are also in the transaction with a high probability. For example, assume that whenever customers in a supermarket buy bread and butter, they will also buy milk. From the supermarket transactions, an association rule such as "*Bread* and *Butter* → *Milk*" will be mined.

Numerous mining approaches have been proposed for association rule mining (Agrawal et al. 1993; Agrawal and Srikant 1994; Cai et al. 1998), with most focusing on binary value transaction data. However, transaction data in real-world applications usually consist of quantitative values. Thus, many fuzzy data mining algorithms have been proposed for deriving fuzzy rules from quantitative transaction databases (Au and Chan 2003; Chan and Au 1998; Dubois et al. 2005; Hong et al. 1999; Hong and Lee 1996; Hong et al. 2003; Kuok et al. 1998; Lee et al. 2008, 2004; Mangalampalli and Pudi 2009, 2010; Ouyang and Huang 2009;

Tajbakhsh et al. 2009; Yue et al. 2000). Since each item has its own utility, utility fuzzy itemset mining has received increasing interest, with many algorithms proposed in recent years (Chan et al. 2003; Chu et al. 2008; Li et al. 2008; Lai et al. 2010; Pillai et al. 2010; Tajbakhsh et al. 2009; Wang et al. 2009). Applications include network alarms (Wu and Li 2011), control (Fazzolari et al. 2013), classification (Kianmehr et al. 2011), and financial analysis (Paranjape-Voditel and Deshpande 2011).

However, there are two common problems with existing fuzzy data mining approaches. First, appropriate minimum support value needs to be set, which is a difficult task. If a large minimum support value is set, only few frequent itemsets will be generated and many potential rules may not be mined. If a small minimum support value is set, a large number of rules will be derived, which cannot be easily used by decision-makers for marketing strategies. The second problem is that some of the derived rules only expose common-sense knowledge, which may not be interesting from a business point of view. For example, if the rule "If *milk* is bought, Then *bread* is bought" is derived with high support and confidence, it will be considered a reliable rule according to the Apriori algorithm (Agrawal and Srikant 1994). However, it may not be valuable for business since the derived rule is common-sense knowledge. There is a problem in existing methods for mining high-utility fuzzy itemsets (HUFIs). Take a derived HUFI {Washington Berry Juice.Middle, Red Wing AA-Size Batteries.Low} from *foodmart* dataset as an example. The derived HUFI means that its utility value is larger than a predefined minimum utility ratio, which also represents the HUFI could reach good profit. However, the derived HUFI does not guarantee its utility value is larger than the utility value of {Washington Berry Juice.Middle, ¬Red Wing AA-Size Batteries.Low} or {¬Washington Berry Juice.Middle, Red Wing AA-Size Batteries.Low}. In other words, such itemset may be misleading.

Recently, Cao (2008, 2010, 2012) and Cao et al. (2010a, b, 2011) suggested the domain-driven data mining concept (D^3M), and combined it with industry knowledge to mine actual and useful information. Under the D^3M concept, Sim et al. (2010) proposed a logic-based approach for deriving coherent rules from binary transactions. In their approach, by using the properties of propositional logic, the relationship between items (i.e., coherent rules) can be derived directly without knowing the appropriate value of the minimum support. The present study proposes a unified interestingness (UI) actionable knowledge discovery (AKD)-based mining framework for high-coherent-utility fuzzy itemset mining. Using the framework, an algorithm for mining high-coherent-utility fuzzy itemsets from quantitative transactions that considers the properties of propositional logic is proposed. The proposed approach first transforms quantitative transactions into fuzzy sets by the predefined membership functions. Then, the utility of each fuzzy itemset is calculated according to the given external utility table. If the value is larger than or equal to the minimum utility ratio, the itemset is a high-utility fuzzy itemset (HUFI). Finally, the contingency tables of each HUFI are calculated and then used for deriving the high-coherent-utility fuzzy itemsets (HCUFIs). Experiments are conducted on the *foodmart* and simulated datasets to show the results of the proposed approach. Comparison results between HCUFI and other approaches first show that the derived itemsets by the proposed algorithm not only can reach better profit than selling them separately, but also can provide fewer but more useful utility itemsets for decision-makers. Efficiency and scalability of HCUFI approach shows that execution times for deriving HCUFIs are decreasing along with the increasing of minimum utility ratios, and it is suitable for mining itemsets from small number of transactions.

The main contributions of the proposed HCUFI algorithm are listed as follows. (1) It is trying to find HCUFIs without minimum support threshold that actually could provide useful utility itemsets for making decisions. (2) Based on the experimental results, take the derived HCUFI "Bought medium amount of High Top Red Delicious Apples, and small amount of Carlson Muenster Cheese, sup = 0.0015, UFI = 13.675" from the *foodmart* dataset as an example. According to the properties of propositional logic, the derived HCUFI shows that the items "High Top Red Delicious Apples" and "Carlson Muenster Cheese" can reach better profit than selling them separately, and may provide users for making a more profitable marketing strategy.

The remainder of this paper is organized as follows. Preliminaries are described in Sect. 2. Related work is reviewed in Sect. 3. The UI-AKD-based mining framework is given in Sect. 4. The proposed HCUFI mining algorithm is described in Sect. 5. Experiments that demonstrate the performance of the proposed algorithm are described in Sect. 6. Conclusions and future works are given in Sect. 7.

2 Preliminaries

In this section, preliminaries are stated. Domain-driven data mining concepts and main issue of minimum support threshold are described in Sects. 2.1 and 2.2, respectively.

2.1 Domain-driven data mining concepts

Domain-driven data mining concepts are proposed by Cao (2008, 2010, 2012) and Cao et al. (2010a, b, 2011), which combined it with industry knowledge to mine useful real information. They emphasize the issues surrounding real-world data mining, and propose the trends from data-centered hidden pattern discovery to domain driven AKD. And, the "actionable" means that the derived knowledge patterns can not only provide important grounds to business decision makers for making appropriate actions, but also deliver expected outcomes to business. The four frameworks for the logical concept of D^3M are described in Cao et al. (2010a). They are post-analysis-based AKD (PA-AKD), UI-based AKD (UI-AKD), combined-mining-based AKD (CM-AKD) and multisource combined-mining-based AKD (MSCM-AKD).

Since the proposed approach is based on the UI-AKD framework, the details are stated. According to Cao et al. (2010b), UI-AKD framework can be formalized as follows:

$$UI\text{-}AKD : DB \xrightarrow{e,\, i(),\, m,\, \Omega_d,\, \Omega_m} \tilde{P}, \tilde{R}$$

where $i()$ can be defined as follows:

$$i() \rightarrow \eta \hat{t}_i() + \varpi \hat{b}(),$$

where weights η and ϖ reflect the interestingness balance between data analysts and domain experts in terms of the business problem, data, environment, and deliverable expectation. $i()$ consists of t_i and b_i, where t_i is technical interestingness and b_i is business interestingness. In some cases, both weights and aggregation can be fuzzy. In other cases, the aggregation may happen in a step-by-step manner. For each step, weights may be differentiated. In practice, the combination of technical interestingness with business expectations may be implemented by various methods. An ideal situation is to generate a single formula $i()$ integrating t_i and b_i, and then to filter patterns accordingly. If such a uniform metric is not available, an alternative way is to calculate t_i and b_i for all patterns, and then rank them in terms of them respectively. Weight-based voting (weights are determined by user) can then be used to aggregate the two ranked lists into a unified pattern set. If there is uncertainty in merging the pattern sets, fuzzy set-based aggregation and ranking may be helpful. The environment (e), domain knowledge (Ω_d), meta-knowledge (Ω_m), technical interestingness, and business interestingness are then merged by method m into a final pattern set ($\tilde{P}$). The patterns are converted into business rules ($\tilde{R}$) as the final deliverables that reflect business preferences and needs. Based on the UI-AKD framework, the details of the proposed HCUFI mining framework and algorithm are presented in Sects. 4 and 5.

2.2 Main issue of minimum support threshold

The main issue of association rule mining approaches is how to define appropriate minimum support and minimum confidence values. A previous study reported that although the appropriate minimum support maybe exist, it is difficult to find (Webb and Zhang 2005). Numerous methods for finding the appropriate minimum support have been proposed (Sim et al. 2010). In general, using different minimum supports to derive association rules maybe result in different mining results. A small minimum support will generate too many frequent itemsets and association rules, which are not easily used for user decision-making. In contrast, a large minimum support will ignore possibly useful itemsets (Sim et al. 2010). Liu et al. (1999) thus proposed an algorithm that uses multiple minimum supports, called minimum item supports (MISs), for mining association rules. Many approaches have been proposed for setting appropriate multiple minimum supports using heuristic methods (Koh et al. 2006; Lin et al. 2002; Yun et al. 2003).

Sim et al. (2010) thus proposed an association rule mining framework for association rule mining without a minimum support threshold. In their approach, using the properties of propositional logic, the relationship between items can be derived directly without knowing the appropriate value of the minimum support. The approach maps the association rules to equivalences. Each mapping from an association rule to an equivalence should satisfy the conditions shown in Table 1.

Table 1 Four conditions for mapping rules to equivalence

Equivalences	$p \equiv q$	$\neg p \equiv \neg q$
Association rules	$X \rightarrow Y$	$\neg X \rightarrow \neg Y$
True or False on association rules	Required conditions	
T	$X \rightarrow Y$	$\neg X \rightarrow \neg Y$
F	$X \rightarrow \neg Y$	$\neg X \rightarrow Y$
F	$\neg X \rightarrow Y$	$X \rightarrow \neg Y$
T	$\neg X \rightarrow \neg Y$	$X \rightarrow Y$

In Table 1, X and Y are two itemsets. An association rule X→Y is mapped to $p \equiv q$ if and only if (1) $X \rightarrow Y$ is true; (2) $\neg X \rightarrow Y$ is false; (3)$X \rightarrow \neg Y$ is false; and (4) $\neg X \rightarrow \neg Y$ is true. When used in multiple transactions, association rules can be mapped to implications as follows: $X \rightarrow Y$ is mapped to an implication $p \rightarrow q$ if and only if (1) Sup(X, Y) > Sup(X, Y); (2) Sup(X, Y) > Sup(¬X, Y); (3) Sup(X, Y) > Sup(X, ¬Y); and (4) Sup(X, Y) > Sup(¬X, ¬Y). In the same way, other association rules mapped to implications based on a comparison between supports can be derived, named *pseudoimplications*. According to the *pseudoimplications*, the derived rules are called coherent rules. The following four conditions must be satisfied for a coherent rule: (1) Sup(X, Y) > Sup(¬X, Y); (2) Sup(X, Y) > Sup(X, ¬Y); (3) Sup(¬X, ¬Y) > Sup(¬X, Y); and (4) Sup(¬X, ¬Y) > Sup(X, ¬Y). These four conditions can also be represented as a contingency table, as shown in Table 2.

By utilizing the coherent rules concept, the present study proposes a fuzzy data mining algorithm that incorporates the four conditions into the mining process for deriving high-coherent-utility fuzzy itemsets from quantitative transactions without a minimum support threshold.

Table 2 Contingency table of a rule

Frequency of co-occurrences	Consequence Y	
	Y	$\neg Y$
Antecedent X		
X	$Q_1 = \text{Sup}(X,Y)$	$Q_2 = \text{Sup}(X, \neg Y)$
$\neg X$	$Q_3 = \text{Sup}(\neg X,Y)$	$Q_4 = \text{Sup}(\neg X, \neg Y)$

3 Related work

This section reviews related work. In Sect. 3.1, binary and fuzzy rule mining approaches are described. Then, utility fuzzy itemset mining approaches are reviewed in Sect. 3.2.

3.1 Binary and fuzzy data mining approaches

The earliest association rule mining was proposed by Agrawal and Srikant (1994). Two criteria are designed to measure the validity of an association rule, namely the support and confidence. It has been used in many fields, such as shopping cart analysis (Agrawal et al. 1993), network intrusion analysis (Tajbakhsh et al. 2009), and stock market analysis (Au and Chan 2003). And, they are designed for mining rules from binary transaction data. However, real-world transactions may have quantitative values, which must be properly handled. Numerous mining algorithms have been proposed for deriving fuzzy rules from quantitative transaction databases (Alcala-Fdez et al. 2010; Dubois et al. 2005; Intan and Yenty 2008; Mangalampalli and Pudi 2009; Matthews et al. 2011; Martin and Shen 2009; Ouyang and Huang 2011; Sathiyapriya et al. 2011; Zhao and Yao 2010).

To obtain appropriate fuzzy support and confidence values, Dubois et al. (2005) investigated techniques to identify and evaluate associations in a relational database that are expressible by fuzzy if-then rules. That paper has exhibited the dependence of the methods for assessing the support for an association by relevance and distribution of the data. In addition to the satisfaction of a confidence threshold, a distribution criterion was proposed to ensure the robustness of the support for an association rule and to mitigate the anomalies that can result from the accumulation of small cardinalities. Martin and Shen (2009) described an approach for deriving association rules between fuzzy categories based on mass assignment theory.

As to different types of fuzzy association rule approaches, Zhao and Yao (2010) presented a general framework for mining fuzzy association rules. Since the traditional association rule mining algorithm is limited to dealing with non-symmetric binary attributes and discrete attributes, Zhao et al.'s framework can be employed for mining distinct types of patterns. Ouyang and Huang (2011) introduced a concept called indirect association rules and proposed a discovery algorithm for mining both direct and indirect fuzzy association rules with multiple minimum supports. Intan and Yenty (2008) proposed an algorithm for multidimensional fuzzy association rule mining from a normalized database by a denormalized table. Matthews et al. (2011) presented a method for mining temporal itemsets by using multi-objective evolutionary search and optimization. Alcala-Fdez et al. (2010) presented a study of three genetic association rule extraction methods to show their effectiveness for mining quantitative association rules. Sathiyapriya et al. (2011) proposed an algorithm for hiding sensitive or important fuzzy association rules. A learning method is first used to derive membership functions automatically for numeric data. Then, a method for preventing the extraction of useful association rules from quantitative data by decreasing the support of the consequence part of the rule is used.

As to the efficient fuzzy association rule mining approaches, Mangalampalli and Pudi (2009) proposed a mining algorithm for mining rules from very large datasets as an alternative to the fuzzy Apriori method. In that approach, two-phase tidlist-style processing is used to compress tidlists and increase the speed of the mining processing.

3.2 Utility fuzzy itemset mining approaches

Utility fuzzy itemset mining is used in various applications, such as, stock market analysis, on-line transaction analysis, and retail analysis. Although traditional association rule mining algorithms generate frequent itemsets and high-confidence rules, these results may not provide useful actionable knowledge for business goals.

Chan et al. (2003) proposed a top-k objective-directed data mining approach. It focuses on a kind of objective-oriented utility-based association mining called top-k high-utility closed-pattern mining that directly supports a given business objective. However, the frequency of a utility itemset may not be a sufficient indicator of interestingness because it only reflects the number of transactions in the data windows of a data stream that contain the itemset. Chu et al. (2008) thus proposed the THUI-Mine algorithm. It can effectively identify temporal high-utility itemsets with few candidate itemsets and reduce execution time. Li et al. (2008) proposed MHUI-BIT and MHUI-TID one-pass algorithms. They use a lexicographical tree-based summary data structure to handle mining high-utility itemsets from data streams within a transaction-sensitive sliding window. Vo et al. (2009) proposed a method for mining high-utility itemsets from vertical distributed databases that uses a WIT-tree technique to scan the local database only once. Pillai et al. (2010) presented a theoretical conceptual approach for temporally weighted itemset utility mining for handling the dynamic utility of itemsets based on cost, profit, or revenue data.

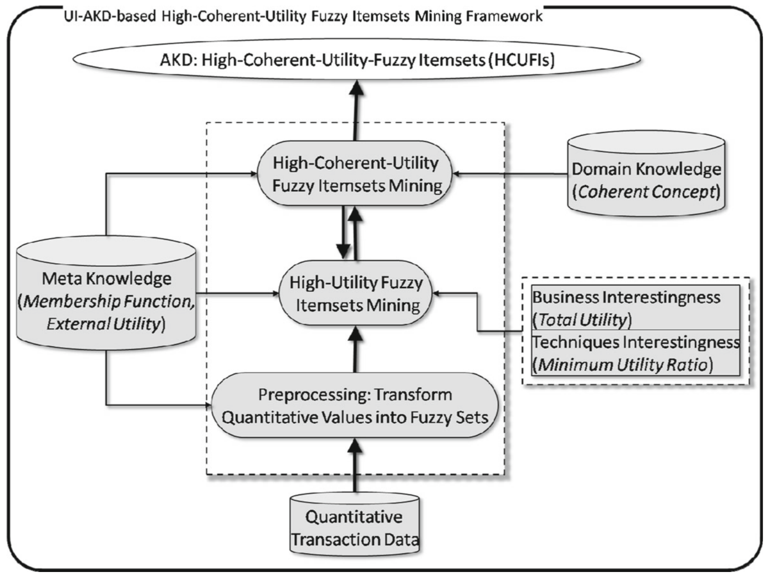

Fig. 1 Proposed UI-AKD-based mining HCUFI mining framework

However, transactions may have quantitative values. Thus, approaches that consider the utility of fuzzy itemsets have been proposed, where the utility means profits. Wang et al. (2009) proposed a method called HUFI-Miner to discover HUFIs from quantitative databases by considering both profits and quantities of items. The quantitative values are first transformed into fuzzy regions. Then, the external utility is used to calculate the utility of each fuzzy itemset. If the utility of a fuzzy itemset is larger than or equal to the minimum utility ratio, the itemset is a HUFI. According to Wang et al. (2009), the definition of HUFI is given as follows.

Definition 1 (*HUFI*) A collection, F_α^k, of high utility fuzzy itemsets with respect to a user specified minimum utility value α, is defined to be the set contains all fuzzy k-itemsets X with $\text{UFI}(X) > \alpha$, where $\text{UFI}(X)$ is the utility value of a fuzzy itemset X in database D.

Then, Lai et al. (2010) proposed FHUI-Mine for mining HUFIs. FHUI-Mine has two phases. In Phase I, the quantitative values are transformed into a fuzzy transaction table according to the given membership functions. Then, the fuzzy transaction table is utilized to find the fuzzy transaction-weighted utility itemsets, which are used to derive high-utility itemsets. In Phase II, the Phase I high-utility itemsets are scanned to find HUFIs. However, those approaches have two problems. The first one is that they need to set minimum support, and it is a hard task. The second one is that the derived itemsets by existing methods may be misleading. This study thus proposes a UI-AKD-based algorithm for mining high-coherent-utility fuzzy itemsets to overcome problems with the properties of propositional logic.

4 Proposed UI-AKD-based mining framework

In this section, based on the D^{3}M concept, propositional logic, and utility fuzzy itemset mining, the UI-AKD-based mining HCUFI mining framework is proposed for mining HCUFIs. The framework of the proposed approach is shown in Fig. 1.

The three main parts in the framework are preprocessing, HUFI mining, and HCUFI mining. In the preprocessing stage, the quantitative values are transformed into fuzzy sets according to the meta knowledge (membership function). Then, in the second stage, the business interestingness and technical interestingness and the meta knowledge are taken into consideration for mining HUFIs, which are total utility, the minimum utility ratio, and external utility, respectively. In the last stage, the domain and meta knowledge are used for deriving actionable knowledge (i.e., HCUFIs). The proposed UI-AKD-based mining framework can be formalized as follows:

$$UI\text{-}AKD: \text{Quantitative Transactions} \xrightarrow{e,\, i(),\, m,\, \Omega_d,\, \Omega_m} \text{HCUFIs}$$

where e is the fuzzy data mining environment, $i()$ consists of the business interestingness and technical interestingness, which are total utility and the minimum utility ratio, respectively. m represents the proposed mining algorithm (for details, please see Sect. 4). Ω_m is meta knowledge, which contains the membership functions and external utility and Ω_d is domain knowledge. The coherence concept is utilized to evaluate the actionable ability of a utility fuzzy itemset. The framework takes the business environment, business and technical interestingness, and meta and domain knowledge into consideration for deriving actionable knowledge patterns using the proposed algorithm m (HCUFI algorithm, see Sect. 4). Note that the UI-AKD framework is one of the general domain-driven data mining framework proposed by Cao et al. for mining actionable knowledge (patterns) (Cao 2010; Cao et al. 2010b). Based on UI-AKD framework, we propose the HCUFI framework for mining HCUFIs. Since domain-driven data mining focus on how to take the objective and subjective interestingness in terms of technical and business goals into consideration for driving actionable knowledge (patterns), based on the HCUFI framework, the HCUFI algorithm is thus proposed to mine HCUFIs.

5 Proposed high-coherent-utility fuzzy itemset mining algorithm

Based on the fuzzy data mining algorithm (Hong et al. 1999) and the coherent rule concept (Sim et al. 2010), the proposed algorithm for mining HCUFIs is described in Sect. 5.1. An example is then given in Sect. 5.2.

5.1 Proposed HCUFI mining algorithm

Firstly, the notations used in the HCUFI algorithm are described as follows:

$D^{(i)}$	The i-th transaction datum, $i = 1$ to n;
I_j	The j-th item, $j = 1$ to m;
$v_j^{(i)}$	The purchased quantity of item I_j in i-th transaction;
R_{jk}	The k-th fuzzy region of item I_j;
$f_{jl}^{(i)}$	$v_j^{(i)}$'s fuzzy membership value in region R_{jl};
$UFI(R_{jl})$	The utility value of fuzzy region R_{jl};
$UFI(R_{jl})$	The weighted external utility of a fuzzy region R_{jl};
$HUFI_k$	The high-utility fuzzy itemset with length k;
$HCUFI_k$	The high-coherent-utility fuzzy itemsets with length k.

In the proposed HCUFI algorithm, STEP 1 to 2 transform quantitative transactions into fuzzy sets using the predefined membership functions. STEP 3 to 4 calculate the utility value of each fuzzy region (R_{jk}). STEP 5 to 6 check utility value of each fuzzy region against the minimum utility value to find the HUFIs with length one ($HUFI_1$). STEP 7 to 9 use the derived $HUFI_1$to mine $HUFI_{k,}$ and then use the mined $HUFI_k$ to obtain the HCUFIs. At last, the derived HCUFIs are outputted at STEP 10. The details of the HCUFI algorithm are stated as follows:

INPUT: A body of n quantitative transaction data, an external utility (EU), a minimum utility ratio α, and a given set of membership functions.

OUTPUT: A set of HCUFIs.

STEP 1: Transform the quantitative value $v_j^{(i)}$ of each transaction datum $D^{(i)}$, $i = 1$ to n, for each item I_j, $j = 1$ to m, into a fuzzy set $f_j^{(i)}$ represented as $\left(\frac{f_{j1}^{(i)}}{R_{j1}} + \frac{f_{j2}^{(i)}}{R_{j2}} + \cdots + \frac{f_{jh}^{(i)}}{R_{jh}}\right)$ using the given membership functions, where R_{jk} is the k-th fuzzy region of item I_j, $v_j^{(i)}$ is the purchased quantity of item I_j in i-th transaction, $f_{jl}^{(i)}$ is $v_j^{(i)}$'s fuzzy membership value in region R_{jl}, and h $(= |I_j|)$ is the number of fuzzy regions for I_j.

STEP 2: For each fuzzy region R_{jl}, calculate its complement value and represent as:

$$\left(\frac{1 - f_{j1}^{(i)}}{R_{j1}} + \frac{1 - f_{j2}^{(i)}}{R_{j2}} + \cdots + \frac{1 - f_{jh}^{(i)}}{R_{jh}}\right).$$

STEP 3: Collect all fuzzy regions into the set A.

STEP 4: Calculate utility of fuzzy itemset value of each fuzzy region R_{jk} in A using:

$$UFI(R_{jl}) = \sum_{i=1}^{n} f_{jl}^{(i)} \times WEU(R_{jl}),$$

where $WEU(R_{jl})$ is the weighted external utility of a fuzzy region R_{jl}, which is defined as:

$$WEU(R_{jl}) = \left[\max\left(f_{jl}^{(1)}, f_{jl}^{(2)}, \ldots, f_{jl}^{(n)}\right)\right] \times EU(I_j),$$

where $\max\left(f_{jl}^{(1)}, f_{jl}^{(2)}, \ldots, f_{jl}^{(n)}\right)$ is the maximum fuzzy value of fuzzy region R_{jl}, and $EU(I_j)$ is the external utility value of item I_j.

STEP 5: Calculate the total utility (*TU*) value of this quantitative transaction data using:

$$TU = \sum_{j=1}^{m} \sum_{l=1}^{k} UFI(R_{jl}).$$

STEP 6: Check each *UFI*(R_{jh}) against the minimum utility value φ (= α * *TU*). If *UFI*(R_{jl}) is larger than or equal to the threshold φ, then put the itemset into the HUFIs with length one as follows:

$$HUFI_1 = \left\{UFI_1 \left| UFI(R_{jl}) \geq \varphi\ ,\ 1 \leq j \leq m,\ 1 \leq l \leq h\right.\right\}$$

STEP 7: If the set $HUFI_1$ is not empty, set $k = 2$, where k is the length of utility fuzzy itemsets. Otherwise, exit the mining procedure.

STEP 8: Do the following substeps to generate HCUFIs with length k:

SUBSTEP 8.1: Join *HUFIs* with length $(k-1)$ to generate the candidate *UFI* with length $k$and put into $CUFI_k$.

SUBSTEP 8.2: Calculate utility value of each itemset S in $CUFI_k$ using:

$$UFI(S) = \sum_{i \in TID(S)} \sum_{R_{jl} \in S} f_{R_{jl}}^{(i)} \times WEU(R_{jl}),$$

where $TID(S)$ represents transactions that contain the itemset S.

SUBSTEP 8.3: Check each $UFI(S)$ to the minimum utility value. If $UFI(S)$ is larger than or equal to the threshold, then put the itemset into HUFIs with length k as follows:

$$HUFI_k = \{UFI_k | UFI(S) \geq \varphi, S \in CUFI_k\}.$$

SUBSTEP 8.4: For each subset of itemset S in $HUFI_k$, calculate the contingency table for antecedent X and consequent Y, where $X \in S$, $Y \in S$, $X = (S - Y)$. The length of Y equals 1. Here, four count values are calculated, namely Q_1:UFI(XY), Q_2: UFI($X \sim$ Y), Q_3: UFI($\sim XY$), and Q_4: UFI($\sim X \sim Y$).

SUBSTEP 8.5: Check whether all subsets of itemset S in $HUFI_k$ meet the four conditions $Q_1 > Q_2$, $Q_1 > Q_3$, $Q_4 > Q_2$, and $Q_4 > Q_3$. If so, add S to the set $HCUFI_k$. If all itemsets in $HCUFI_k$ are checked, let $HCUFI_{all} = HCUFI_{all} \cup HCUFI_k$ and go to SUBSTEP 8.6. Otherwise, go to SUBSTEP 8.4.

SUBSTEP 8.6: Check whether the $HCUFI_k$ set is empty. If so, go to STEP 10. Otherwise, go to STEP 9.

STEP 9: Set $k = k + 1$ and go to STEP 8.

STEP 10: Output the derived $HCUFI_{all}$.

5.2 An example

In this section, an example is given to illustrate the HCUFI algorithm. This is a simple example to show how it can be used to mine HCUFIs from quantitative transaction data. There are four items in a transaction database: milk, bread, cookies, and beverage. The dataset includes the six transactions shown in Table 3.

The fuzzy membership functions for the four items are shown in Fig. 2.

The external utility values for the four items are shown in Table 4.

Table 3 Six transactions in this example

TID	Items
$T1$	(milk, 10); (bread, 10); (cookies, 7), (beverage, 7)
$T2$	(milk, 12); (bread, 14); (cookies, 12)
$T3$	(bread, 2); (cookies, 12)
$T4$	(milk, 2); (bread, 4); (cookies, 5)
$T5$	(milk, 9); (bread, 9)
$T6$	(milk, 5); (beverage, 12)

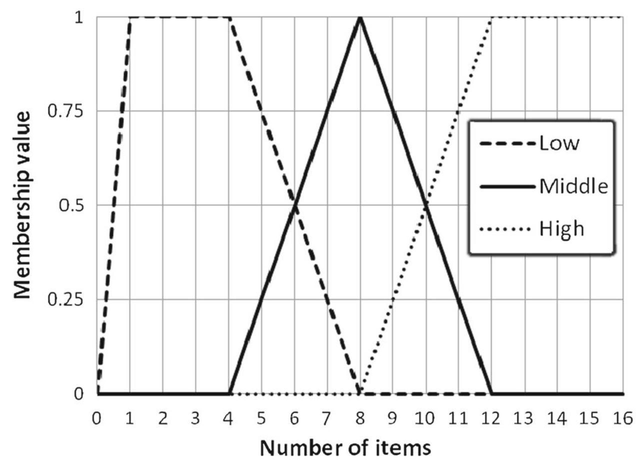

Fig. 2 Membership functions used in this example

Table 4 The external utility values for items

Item	External utility value	Item	External utility value
milk	5	*cookies*	1
bread	3	*beverage*	7

Table 5 Fuzzy sets transformed from the data in Table 3

TID	Fuzzy set
T1	$\left(\frac{0.5}{milk.Middle} + \frac{0.5}{milk.High}\right)\left(\frac{0.5}{bread.Middle} + \frac{0.5}{bread.High}\right)$ $\left(\frac{0.25}{cookies.Low} + \frac{0.75}{cookies.Middle}\right)\left(\frac{0.25}{beverage.Low} + \frac{0.75}{beverage.Middle}\right)$
T2	$\left(\frac{1}{milk.High}\right)\left(\frac{1}{bread.High}\right)\left(\frac{1}{cookies.High}\right)$
T3	$\left(\frac{1}{bread.Low}\right)\left(\frac{1}{cookies.High}\right)$
T4	$\left(\frac{1}{milk.Low}\right)\left(\frac{1}{bread.Low}\right)\left(\frac{0.75}{cookies.Low} + \frac{0.25}{cookies.Middle}\right)$
T5	$\left(\frac{0.75}{milk.Middle} + \frac{0.25}{milk.High}\right)\left(\frac{0.75}{bread.Middle} + \frac{0.25}{bread.High}\right)$
T6	$\left(\frac{0.75}{milk.Low} + \frac{0.25}{milk.Middle}\right)\left(\frac{1}{beverage.High}\right)$

Table 4 shows that the external utility values of milk, bread, cookies, and beverage are 5, 3, 1, 7, respectively. The minimum utility ratio α is set at 10 %. Each item has three fuzzy regions: *Low*, *Middle*, and *High*. Thus, three fuzzy membership values are produced for each item according to the predefined membership functions. For the transaction data in Table 3, the HCUFI algorithm proceeds as follows.

STEP 1: The quantitative value of each transaction datum is transformed into a fuzzy set according the predefined membership functions. Take the second item in transaction $T5$ using the membership functions as an example. The amount "9" of item *bread* is then converted into the fuzzy set $\left(\frac{0.75}{bread.Middle} + \frac{0.25}{bread.High}\right)$ using the given membership functions. The results for all items are shown in Table 5, where the notation *item.term* is called a fuzzy region.

Table 6 UFI value of each fuzzy region in A

Fuzzy region	*UFI*	Fuzzy region	*UFI*
milk.Low	8.75	*cookies.Low*	1
milk.Middle	60	*cookies.Middle*	8
milk.High	105	*cookies.High*	24
bread.Low	6	*beverage.Low*	1.75
bread.Middle	30	*beverage.Middle*	42
bread.High	63	*beverage.High*	84

Table 7 UFI values of all itemsets in CUFI_2

ID	Itemset *S*	$UFI(S)$
1	(*milk.Middle, bread.High*)	77
2	(*milk.Middle, beverage.High*)	94
3	(*milk.High, bread.High*)	168
4	(*milk.High, beverage.High*)	0
5	(*bread.High, beverage.High*)	0

Table 8 Contingency table for (milk.High, bread.High)

Frequency of co-occurrences	Consequence *Y*	
	$Y = \{bread.High\}$	$\sim Y = \sim\{bread.High\}$
Antecedent *X*		
$X = \{milk.High\}$	168 (Q_1)	90 (Q_2)
$\sim X = \sim\{milk.High\}$	102 (Q_3)	408 (Q_4)

STEP 2: The fuzzy value is transformed into a complement fuzzy set. Take the second item in transaction $T5$ using the membership functions as an example. The fuzzy region $(\frac{0.75}{bread.Middle} + \frac{0.25}{bread.High})$ of item *bread* is then converted into the complement fuzzy set $(\frac{1}{bread.Low} + \frac{0.25}{bread.Middle} + \frac{0.75}{bread.High})$. The results for all items are calculated in the same way.

STEP 3: All fuzzy regions are then collected into set A, which is $A = \{$*milk.Low, milk.Middle, milk.High, bread.Low, bread.Middle, bread.High, cookies.Low, cookies.Middle, cookies.High, beverage.Low, beverage.Middle, beverage.High*$\}$.

STEP 4: The utility of each fuzzy itemset is calculated. Take fuzzy region "*milk.Middle*" as an example. UFI(*milk.Middle*) is 60 $(=(0.5+0+0+0+0.75+0.25) \times (8 \times 5))$. The results of other fuzzy regions are shown in Table 6.

STEP 5: The total utility of the given transactions is then calculated. The *TU* of this example is 433.5 (= UFI(*milk.Low*) + UFI(*milk.Middle*) + UFI(*milk.High*) + UFI(*bread.Low*) + UFI(*bread.Middle*) + UFI(*bread.High*) + UFI(*cookies.Low*) + UFI(*cookies.Middle*) + UFI(*cookies.High*) + UFI(*beverage.Low*) + UFI(*beverage.Middle*) + UFI(*beverage.High*) = $1.75 \times 1 \times 5 + 1.5 \times 8 \times 5 + 1.75 \times 12 \times 5 + 2 \times 1 \times 3 + 1.25 \times 8 \times 3 + 1.75 \times 12 \times 3 + 1 \times 1 \times 1 + 1 \times 8 \times 1 + 2 \times 12 \times 1 + 0.25 \times 1 \times 7 + 0.75 \times 8 \times 7 + 1 \times 12 \times 7$).

STEP 6: Since the minimum utility ratio is set at 10 %, the minimum utility value is calculated as 43.35 (= 10 % × 433.5). The set $HUFI_1$ thus contains *milk.Middle*, *milk.High*, *bread.High*, and *beverage.High*.

STEP 7: Since $HUFI_1$ is not empty, the parameter k is set at 2.

STEP 8: The HCUFIs are derived by the following substeps:

SUBSTEP 8.1: In this step, the itemsets in $HUFI_1$ are joined to generate candidate *UFI* with length 2: $CUFI_2 = \{$(*milk.Middle, bread.High*), (*milk.Middle, beverage.High*), (*milk.High, bread.High*), (*milk.High, beverage.High*), (*bread.High, beverage.High*)$\}$.

SUBSTEP 8.2: The utility of each fuzzy itemset is calculated. Take candidate utility fuzzy itemset "(*milk.High, bread.High*)" as an example. Since the fuzzy itemset appears in T1, T2, and T5, $UFI_{(milk.High,bread.High)}$ is 168 (= 105 + 63). The UFI values of other itemsets are shown in Table 7.

SUBSTEP 8.3: Each utility fuzzy itemset is checked against the minimum utility ratio. With the minimum utility ratio set at 0.1, since the UFI value of itemset (*milk.High*, *bread.High*) is larger than 43.35 (= 0.1 × 433.5), the itemset is added to $HUFI_2$.

SUBSTEP 8.4: The subset of each fuzzy itemset is used to calculate its contingency table. Take the HUFI "(*milk.High*, *bread.High*)" as an example. The contingency tables of (*milk.High* → *bread.High*) and (*bread.High* → *milk.High*) should be calculated. Take (*milk.High* → *bread.High*) as an example. The consequence part Y is *bread.High* and the antecedent part X is *milk.High*. Four UFI values need to be calculated. They are Q_1: UFI(*milk.High*, *bread.High*), Q_2: UFI(*milk.High*, ∼*bread.High*), Q_3: UFI(∼ *milk.High*, *bread.High*), and Q_4: UFI(∼*milk.High*, ∼*bread.High*). The UFI value of each itemset can be calculated by equation defined in Sect. 5.1 (Step 8.2). The contingency table for itemset (*milk.High*, *bread.High*) is show in Table 8.

SUBSTEP 8.5 to: From Table 8, (*milk.High*, *bread.High*) meets the four conditions, with 168 > 90, 168 > 102, 408 > 90, and 408 > 102. Thus, (*milk.High*, *bread.High*) is put into *HCUFI*$_2$. In the same way, other HCUFIs can be derived. Since a HCUFI (*milk.High*, *bread.High*) is generated, it go to STEP 9.

STEP 9 to 10: The parameter k is set at 3, it go to STEP 8. In this example, no more HCUFIs can be generated. Thus, only one HCUFI, (*milk.High*, *bread.High*), is outputted.

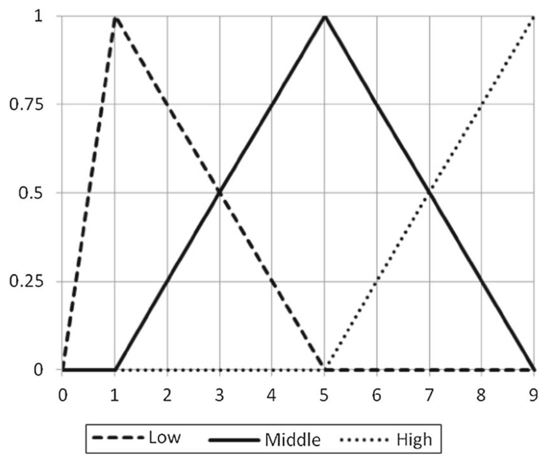

Fig. 3 Membership functions used in the experiments

6 Experiment results

In this section, experimental results of the proposed approach are presented. The experiments were implemented in Java on a personal computer with an Intel Core i7 2.93-GHz CPU and 4 GB RAM. Two datasets are described in Sect. 6.1. The comparison results between HCUFI and other approaches are given in Sect. 6.2. The efficiency and scalability of HCUFI Approach are stated in Sect. 6.3.

6.1 Dataset descriptions

The *foodmart* database (Microsoft Corporation) was used to evaluate the performance of the algorithms. The database is from Microsoft SQL Server 2000. The dataset contains 21,556 transactions with 1,600 different items. In the experiments, 1,000 transactions were used to evaluate the proposed approach. The membership functions are given in Fig. 3. Three fuzzy sets were used in the experiments, namely *Low*, *Middle*, and *High*.

The simulated dataset was generated from the IBM data generator (IBM Quest Data Mining Project 1996). We also developed a simulation model, which was similar to that used in Liu et al. (2005), to generate the quantities of the items in the transactions. Each quantity ranged among 1 to 5 according to the way in Liu et al. (2005). The utility value of each item is generated randomly among 1 to 5. The adopted parameters, T, I, N and D, represented the average length of items per transaction, the average length of maximal potentially frequent itemsets, the total number of different items, and the total number of transactions, respectively. These parameters T, I, N and D were set at 10, 4, 4k and 10k, respectively. In the experiments, 1,000 and 10,000 transactions were used to evaluate the proposed approach.

Table 9 Comparison between HCUFI and FAR approaches

	HCUFI algorithm	FAR algorithm
Number of itemsets (length = 2)	21	3,620
Number of itemsets (length = 3)	1	145
The average support	0.000841	0.000416

6.2 Comparison results between HCUFI and other approaches

Experiments were conducted to compare the derived itemsets obtained using the proposed HCUFI and the original fuzzy association rule (FAR) mining approach (Hong et al. 1999). The minimum support and minimum confidence values of FAR were set at 0.000066 and 0.1, respectively. The minimum utility ratio of HCUFI was set at 0.0001. The results are shown in Table 9.

Table 9 shows that the number of derived itemsets obtained using the HCUFI algorithm is smaller than that obtained using the FAR algorithm. This result is reasonable because the goal of the HCUFI algorithm is to find actionable knowledge patterns. In addition, the average support of derived itemsets is larger than that obtained using the FAR algorithm, which means that the derived itemsets obtained by HCUFI approach are more effective. In order to show the merits of the proposed approach more clearly, a derived HCUFI is given below.

$$S = \{671.M, 62.M\},\ \text{sup} = 0.001,\ Q_1 = 33.75,$$
$$Q_2 = 28.975,\ Q_3 = 0,\ Q_4 = 33701.92$$

From the itemset S, since Q_1 is larger than Q_2 and Q_3, it means that these two products are purchased together and can reach a certain profit (= 33.75). In other words, the profit is higher than that obtained selling them separately. In addition, although the support value of the pattern is small, it may be useful in terms of its utility value. However, this pattern is pruned by the FAR algorithm if the minimum support is set to larger than 0.000845. Thus, the proposed algorithm provides fewer but more useful utility itemsets for decision-makers.

Since the *foodmart* dataset has taxonomy, we further choose items in the food and non-consumable categories and for the following experiments. Tables 10 and 11 shows the comparison results of the derived itemsets between the HUFI-Miner (Wang et al. 2009) and HCUFI approach of food and non-consumable categories, respectively.

In Table 10, although the HUFI-Miner could derive larger number of itemsets than the proposed approach, the average support of those itemsets are smaller than that by HCUFI approach. A derived high-coherent-utility fuzzy itemset (length = 2) is given as follows:

Table 10 Comparison results of HUFI-Miner and HCUFI approach in food category

	HUFI-Miner	HCUFI approach
Number of itemsets (length = 2)	7,780	13
Number of itemsets (length = 3)	4	3
Number of itemsets (length = 4)	1	1
Total number of itemsets	7,785	17
The average support	0.000432	0.000852

Table 11 Comparison results of HUFI-Miner and HCUFI approach in non-consumable category

	HUFI-Miner	HCUFI approach
Number of itemsets (length = 2)	690	4
Number of itemsets (length = 3)	1	0
Total number of itemsets	691	4
The average support	0.000420	0.000875

Table 12 The number of HCUFIs and execution times with different minimum utility ratios of the foodmart dataset

α	0.01 %	0.01 %	0.03 %	0.04 %	0.05 %
Number of HCUFIs	22	14	11	7	6
Execution time (s)	1,785	914	542	345	221

Table 13 The number of HCUFIs and execution times with different minimum utility ratios of the simulated dataset

α	0.01 %	0.01 %	0.03 %	0.04 %	0.05 %
Number of HCUFIs	53	34	20	16	13
Execution time (s)	636	424	245	195	140

Table 14 The execution times of HCUFI approach on different transaction sizes

Transaction sizes	$1K$	$3K$	$5K$	$7K$	$10K$
Execution time (s)	8,778	26,161	45,939	69,219	98,797

"Bought medium amount of High Top Red Delicious Apples, and small amount of Carlson Muenster Cheese, sup = 0.0015, UFI = 13.675"

According to the properties of propositional logic, this itemset indicates medium amount of High Top Red Delicious Apples and small amount of Carlson Muenster Cheese can reach better profit than selling them separately, and could provide users for making a more profitable marketing strategy. From Table 11, we can observe the results derived from non-consumable category are similar to Table 10. The derived high-coherent-utility fuzzy itemset is shown as follows:

"Bought medium amount of Black Tie Eyeglass Screwdriver, and medium amount of High Quality Scissors, sup = 0.001, UFI = 19.55"

The derived itemset shows that "Black Tie Eyeglass Screwdriver" and "High Quality Scissors" are bought together with utility value is equal to 19.55. They also mean that they can reach better profit than selling them separately. These derived HCUFIs could be used as the promotion itemsets when a retailer has promotional activities.

6.3 Efficiency and scalability of HCUFI approach

The experiments were made to show the relationship between the number of derived HCUFIs, minimum utility ratios and the execution times on the *foodmart* and simulated datasets. In order to speed up the execution time of the proposed approach, the multi-thread technique is used in the follow experiments. The experiments are conducted in a single run. Results of *foodmart* and simulated datasets are shown in Tables 12 and 13, respectively.

Table 12 shows that the number of derived HCUFIs is decreasing along with the increasing of minimum utility ratios. The execution times of the proposed approach for deriving HCUFIs are also decreasing along with the increasing of minimum utility ratios. Table 13 shows that the number of derived HCUFIs and the execution times of the proposed approach are also decreasing along with the increasing of the minimum utility ratios.

At last, the experiments were made to show the execution times of HCUFI approach on different transaction sizes of simulated datasets. When the minimum utility ratio was set at 0, the results are shown in Table 14.

Table 14 shows that the execution times of the HCUFI approach are increasing linearly along with the increasing of transaction sizes. From the results, we can observe that the proposed approach is time-consuming. This results are reasonable because the minimum utility ratio was set at 0 such that all HUFIs were generated. If the minimum utility ratio was set larger than 0.01 %, then the execution time was decreased (see Table 13). Thus, the HCUFI approach is suitable for mining itemsets from small number of transactions when the minimum utility ratio was set larger than zero.

In the following, the time complexity of the proposed approach is analyzed. The time-consuming parts of the HCUFI approach are Steps 1 to 7 for deriving HUFIs and Steps 8 to 9 for mining HCUFIs from derived HUFIs. Thus, assume the average transaction width is w, the time complexity of HCUFI approach is $O(\sum_{k=2}^{w}(|HUFI_k|^2 * numberTran * w + |HUFI_k| * |S| * numberTran * w))$, where $|HUFI_k|$ means number of high-utility fuzzy itemsets with length k in $HUFI_k$, $|S|$ is the number of fuzzy regions in the itemset S, and $S \in HUFI_k$, and *numberTran* is the transaction size. Thus, $|HUFI_k|^2$ is the time for generating candidate HUFIs. For each candidate HUFI, *numberTran* * *w* is needed to check it is a HUFI or not. Then, for each HUFI in $HUFI_k$, $|S|$ * *numberTran* * *w* is needed to check it is a HCUFI or not. According to the time complexity of the HCUFI approach, it is easily to know that the execution time of the HCUFI approach is larger than existing approaches, e.g., FAR (Hong et al. 1999), HUFI-Miner (Wang et al. 2009). However, the HCUFI approach has two advantages. It can not only mine HCUFIs without minimum support, but also provide fewer but more useful utility itemsets for decision-makers.

7 Conclusion and future works

Finding actionable knowledge patterns is an important issue in data mining. This study thus proposed a UI-AKD-based mining framework for deriving HCUFIs according to the D^3M concept. The framework takes the business environment, business and technical interestingness, and meta and domain knowledge into consideration for deriving actionable knowledge patterns. Based on the proposed framework, an HCUFI mining algorithm was proposed for quantitative transactions. Quantitative transactions are first converted into fuzzy sets. Then, the utility of each fuzzy itemset is calculated according to the given external utility table. If the value is larger than or equal to the minimum utility ratio, the itemset is a HUFI. Finally, contingency tables are derived and used for checking whether a HUFI satisfies specific criteria. If so, it is a HCUFI. Experiments were conducted on the *foodmart* and simulated datasets to show the merits of the proposed approach. The main contribution of this work is that the proposed approach can derive actionable knowledge patterns effectively without setting the minimum support. For example, a derived HCUFI like "{A.L, B.H}" means that when item A.L and item B.L are bought together, a higher profit can be obtained compared to that obtained when selling the items separately. Using this information, a decision-maker can establish a more profitable marketing strategy.

In the future, to speed up the mining processing, some techniques can be used for achieving this purpose, e.g., parallel processing (Fakhrahmad and Dastghaibyfard 2011), projection technique (Lan et al. 2012), or sampling techniques (Sarwar et al. 2000). Besides, the membership functions may also have a critical impact on the final results. Some GA-based approaches for deriving appropriate membership functions from the transactions could also be used to enhance the proposed approach, e.g. the GA-based 2-tuples linguistic representation model (Alcala-Fdez et al. 2009), and a GA-based approach for derive a predefined number of membership functions to obtain the maximum profit within a user specified interval of minimum supports (Kaya and Alhajj 2005).

Acknowledgments This research was supported by the National Science Council of the Republic of China under grant NSC 102-2221-E-032 -056.

References

Alcala-Fdez J, Alcala R, Gacto MJ, Herrera F (2009) Learning the membership function contexts for mining fuzzy association rules by using genetic algorithms. Fuzzy Sets Syst 160(7):905–921

Agrawal R, Imielinksi T, Swami A (1993) Mining association rules between sets of items in large database. The ACM Special Interest Group on Management of Data Conference, pp 1–10

Alcala-Fdez J, Flugy-Pape N, Bonarini A, Herrera F (2010) Analysis of the effectiveness of the genetic algorithms based on extraction of association rules. J Fund Inf Intell Data Anal Granular Comput 98(1)

Agrawal R, Srikant R (1994) Fast algorithms for mining association rules. International conference on very large data, bases pp 487–499

Au WH, Chan KCC (2003) Mining fuzzy association rules in a bank-account database. IEEE Trans Fuzzy Syst 11(2):238–248

Chan KCC, Au WH (1998) An effective algorithm for discovering fuzzy rules in relational databases. IEEE Int Conf Fuzzy Syst 2: 1314–1319

Cai CH, Fu WC, Cheng CH, Kwong WW (1998) Mining association rules with weighted items. The international database engineering and applications symposium, pp 68–77

Cao L (2008) Domain driven data mining (D^3M). IEEE International conference on data mining workshops, pp 74–76

Cao L (2010a) Domain-driven data mining: challenges and prospects. IEEE Trans Knowl Data Eng 22(6):755–769

Cao L, Yu PS, Zhang C, Zhao Y (2010a) Domain driven data mining. Springer, Berlin

Cao L, Zhao Y, Zhang H, Luo D, Zhang C, Park EK (2010b) Flexible frameworks for actionable knowledge discovery. IEEE Trans Knowl Data Eng 22(9):1299–1312

Cao L (2012) Actionable knowledge discovery and delivery. WIREs Data Mining Knowl Discov 2(2):149–163

Cao L, Zhang H, Zhao Y, Luo D, Zhang C (2011) Combined mining: discovering informative knowledge in complex data. IEEE Trans Syst Man Cybern Part B 41(3):699–712

Chan R, Yang Q, Shen Y (2003) Mining high utility itemsets. The IEEE international conference on data mining, pp 19–26

Chu C, Tseng VS, Liang T (2008) An efficient algorithm for mining temporal high utility itemsets from data streams. J Syst Softw 81(7)

Dubois D, Prade H, Sudkamp T (2005) On the representation, measurement, and discovery of fuzzy associations. IEEE Trans Fuzzy Syst 13(2):250–262

Fazzolari M, Alcal'a R, Nojima Y, Ishibuchi H, Herrera F (2013) A review of the application of multi-objective evolutionary fuzzy systems: current status and further directions. IEEE Trans Fuzzy Syst 21(1):45–65

Fakhrahmad SM, Dastghaibyfard GH (2011) An efficient frequent pattern mining method and its parallelization in transactional databases. J Inf Sci Eng 27(2):511–525

Hong TP, Kuo CS, Chi SC (1999) Mining association rules from quantitative data. Intell Data Anal 3(5):363–376

Hong TP, Lee CY (1996) Induction of fuzzy rules and membership functions from training examples. Fuzzy Sets Syst 84:33–47

Hong TP, Lin KY, Chien BC (2003) Mining fuzzy multiple-level association rules from quantitative data. Appl Intell 18(1):79–90

IBM Quest Data Mining Project (1996) Quest synthetic data generation code. http://www.almaden.ibm.com/cs/quest/syndata.html

Intan R, Yenty O (2008) Mining multidimensional fuzzy association rules from a normalized database. International conference on convergence and hybrid information technology, pp 425–432

Kaya M, Alhajj R (2005) Genetic algorithm based framework for mining fuzzy association rules. Fuzzy Sets Syst 152(3):587–601

Kuok CM, Fu AWC, Wong MH (1998) Mining fuzzy association rules in databases. ACM SIGMOD Record 27(1):41–46

Koh YS, Rountree N, O'Keefe RA (2006) Finding non-coincidental sporadic rules using apriori-inverse. Int J Data Warehousing Mining 2:38–54

Kianmehr K, Kaya M, ElSheikh AM, Jida J, Alhajj R (2011) Fuzzy association rule mining framework and its application to effective fuzzy associative classification. WIREs Data Mining Knowl Discov 1(6):477–495

Li H, Huang H, Chen Y, Liu Y, Lee S (2008) Fast and memory efficient mining of high utility itemsets in data streams. IEEE international conference on data mining, pp 881–886

Lai C, Chung P, Tseng VS (2010) A novel algorithm for mining fuzzy high utility itemsets. Inf Control Int J Innov Comput 6(10)

Lee YC, Hong TP, Wang TC (2008) Multi-level fuzzy mining with multiple minimum supports. Expert Syst Appl 34(1):459–468

Lee YC, Hong TP, Lin WY (2004) Mining fuzzy association rules with multiple minimum supports using maximum constraints. Lecture Notes Comput Sci 3214:1283–1290

Liu B, Hsu W, Ma Y (1999) Mining association rules with multiple minimum supports. ACM SIGKDD conference on knowledge discovery and data mining, pp 337–341

Lan GC, Hong TP (2012) A projection-based approach for discovering high average-utility itemsets. J Inf Sci Eng 28:193–209

Liu Y, Liao W, Choudhary A (2005) A fast high utility itemsets mining algorithm. The utility-based data mining, workshop, pp 90–99

Lin WY, Tseng MC, Su JH (2002) A confidence-lift support specification for interesting associations mining. The Pacific-Asia conference on advances in knowledge discovery and data mining, pp 148–158

Microsoft Corporation, Example Database FoodMart of Microsoft Analysis Services

Mangalampalli A, Pudi V (2009) Fuzzy association rule mining algorithm for fast and efficient performance on very large datasets. IEEE international conference on fuzzy systems, pp 1163–1168

Mangalampalli A, Pudi V (2010) FPrep: fuzzy clustering driven efficient automated pre-processing for fuzzy association rule mining. IEEE international conference on fuzzy systems, pp 1–8

Matthews SG, Gongora MA, Hopgood AA (2011) Evolving temporal fuzzy itemsets from quantitative data with a multi-objective evolutionary algorithm. IEEE international workshop on genetic and evolutionary fuzzy systems, pp 9–16

Martin T, Shen Y (2009) Fuzzy association rules in soft conceptual hierarchies. Annual meeting of the North American Fuzzy Information Processing Society, pp 1–6

Ouyang W, Huang Q (2009) Mining direct and indirect weighted fuzzy association rules in large transaction databases. Int Conf Fuzzy Syst Knowl Discov 3:128–132

Ouyang W, Huang Q (2011) Mining direct and indirect fuzzy association rules with multiple minimum supports in large transaction databases. Int Conf Fuzzy Syst Knowl Discov 2:947–951

Pillai J, Vyas OP, Soni S, Muyeba M (2010) A conceptual approach to temporal weighted item set utility mining. Int J Comput Appl 1(28)

Paranjape-Voditel P, Deshpande U (2011) An association rule mining based stock market recommender system. International conference on emerging applications of information technology, pp 21–24

Sarwar B, Karypis G, Konstan J, Riedl J (2000) Analysis of recommendation algorithms for e-commerce. ACM conference on electronic commerce, pp 158–167

Sathiyapriya K, Sadasivam GS, Celin N (2011) A new method for preserving privacy in quantitative association rules using DSR approach with automated generation of membership function. World congress on information and communication technologies, pp 148–153

Sim ATH, Indrawan M, Zutshi S, Srinivasan B (2010) Logic-based pattern discovery. IEEE Trans Knowl Data Eng 22(6):798–811

Tajbakhsh A, Rahmati M, Mirzaei A (2009) Intrusion detection using fuzzy association rules. Appl Soft Comput 9(2):462–469

Vo B, Nguyen H, Le B (2009) Mining high utility itemsets from vertical distributed databases. International conference on computing and communication technologies, pp 1–4

Wang C, Chen S, Huang Y (2009) A fuzzy approach for mining high utility quantitative itemsets. IEEE international conference on fuzzy systems, pp 1909–1913

Webb GI, Zhang S (2005) k-optimal rule discovery. Data Mining Knowl Discov 10(1):39–79

Yun H, Ha D, Hwang B, Ryu KH (2003) Mining association rules on significant rare data using relative support. J Syst Softw 67:181–191

Yue S, Tsang E, Yeung D, Shi D (2000) Mining fuzzy association rules with weighted items. The IEEE international conference on systems, man and cybernetics, pp 1906–1911

Wu J, Li X (2011) Mining multidimensional fuzzy association rules of alarms in communication networks. International conference on computer science and service, system, pp 2326–2330

Zhao J, Yao L (2010) A general framework for fuzzy data mining. International conference on computational intelligence and software engineering, pp 1–3