Journal of King Saud University – Computer and Information Sciences (2014) 26, 99–109

King Saud University

Journal of King Saud University – Computer and Information Sciences

www.ksu.edu.sa
www.sciencedirect.com

ORIGINAL ARTICLE

Unexpected rules using a conceptual distance based on fuzzy ontology

Mohamed Said Hamani [a], Ramdane Maamri [b,*], Yacine Kissoum [c], Maamar Sedrati [d]

[a] Lire Laboratory, University of M'sila, Algeria
[b] Lire Laboratory, University of Constantine 2, Algeria
[c] Lire Laboratory, University of Skikda, Algeria
[d] University of Hadj Lakhdar Batna, Algeria

Received 31 July 2012; revised 20 April 2013; accepted 3 June 2013
Available online 18 June 2013

KEYWORDS
Fuzzy ontology;
Unexpectedness;
Association rule;
Domain knowledge;
Interestingness;
Conceptual distance

Abstract One of the major drawbacks of data mining methods is that they generate a notably large number of rules that are often obvious or useless or, occasionally, out of the user's interest. To address such drawbacks, we propose in this paper an approach that detects a set of unexpected rules in a discovered association rule set. Generally speaking, the proposed approach investigates the discovered association rules using the user's domain knowledge, which is represented by a fuzzy domain ontology. Next, we rank the discovered rules according to the conceptual distances of the rules.

1. Introduction

Knowledge discovery in data mining has been defined in Fayyad et al. (1996) as the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns from data. Association rule algorithms (Agrawal et al., 1993) are rule-discovery methods that discover patterns in the form of IF-THEN rules. It has been noticed that most of the algorithms that perform data mining generate a large number of rules that are valid but obvious or not interesting to the user (Liu and Hsu, 1996; Piatetsky-Shapiro, 1996; Piatetsky-Shapiro and Matheus, 1991; Silberschatz and Tuzhilin, 1996). To address this issue, most of the approaches to knowledge discovery use objective measures of interestingness for the evaluation of the discovered rules, such as confidence and support measures (Agrawal et al., 1993). These approaches capture the statistical strength of a pattern. The interestingness of a rule is essentially subjective (Liu and Hsu, 1996; Piatetsky-Shapiro and Matheus, 1991; Silberschatz and Tuzhilin, 1996; Klemettinen et al., 1994). Subjective measures of interestingness, such as unexpectedness (Zimmermann, 2001; Asa and Mangano, 1995; Uthurusamy et al., 1991), assume that the interestingness of a pattern depends on the decision-maker and does not solely depend on the statistical strength of the pattern. Although objective measures are

* Corresponding author. Tel.: +213 661693966.
E-mail addresses: saidhamani@hotmail.com (M.S. Hamani), rmaamri@ yahoo.fr, rmaamri@umc.edu.dz (R. Maamri), kissoumyacine @yahoo.fr (Y. Kissoum), msedrati2006@yahoo.fr (M. Sedrati).
Peer review under responsibility of King Saud University.

ELSEVIER Production and hosting by Elsevier

1319-1578
http://dx.doi.org/10.1016/j.jksuci.2013.06.001

useful, they are insufficient in the determination of the interestingness of the rules. One way to address this problem is by focusing on discovering unexpected patterns (Liu and Hsu, 1996; Silberschatz and Tuzhilin, 1996; Liu et al., 1997; Padmanabhan and Tuzhilin, 1998; Padmanabhan and Tuzhilin, 1999; Silberschatz and Tuzhilin, 1995), where the unexpectedness of the discovered patterns is usually defined relative to a system of prior expectations.

Moreover, ontology represents knowledge. Ontology is organized as a DAG (Directed Acyclic Graph) hierarchy. Ontologies allow domain knowledge to be represented explicitly and formally in such a way that it can be shared among human and computer systems. Unfortunately, knowledge about a system can contain ambiguity and vagueness. For this reason, fuzzy ontologies have been used to address such fuzzy knowledge (xxx, 1291), where the concepts are related to each other in the ontology with a degree of membership μ ($0 \leqslant \mu \leqslant 1$). In this paper, we propose a new approach that adds intelligence and autonomy for ranking rules according to their conceptual distance (the distance between the antecedent and the consequent of the rule) relative to the hierarchy. In other words, highly related concepts are grouped together in the hierarchy. The more distant the concepts are, the less they are related to each other. For concepts that are part of the definition of a rule, the less the concepts are related to each other, the more the rule is surprising and therefore interesting. With such a ranking method, a user can check fewer rules on the top of the list to extract the most pertinent ones.

1.1. Association Rules

Association rule mining finds interesting associations and/or correlation relationships among a large set of data items. Association rules show attribute value conditions that occur frequently together in a given dataset. A typical and widely used example of association rule mining is Market Basket Analysis (http://www.resample.com/xl). In market basket analysis, customers' buying habits are analyzed to find associations between different items that customers place in their shopping cart. Two different items, 'a' and 'b', in an item set are assumed to have a relation if they are purchased together in the same transaction. The more those two items are purchased together in the same transaction, the more they have a stronger relation. The discovery of such associations can help retailers develop marketing strategies by gaining an insight into which items are frequently purchased together by customers. Association rules provide information of this type in the form of "if-then" statements. These rules are computed from the data and, unlike the if-then rules of logic, the association rules are probabilistic in nature (http://www.resample.com/xl). Objective measures such as support and confidence are often used to evaluate the interestingness of the association rules.

The support is simply the number of transactions that include all of the items in the antecedent and consequent parts of the rule. The support is sometimes expressed as a percentage of the total number of records in the database.

The confidence is the ratio of the number of transactions that include all of the items in the consequent as well as the antecedent (namely, the support) to the number of transactions that include all of the items in the antecedent.

One way to think of support is that it is the probability that a randomly selected transaction from the database will contain all of the items in the antecedent and consequent, whereas the confidence is the conditional probability that a randomly selected transaction will include all of the items in the consequent given that the transaction includes all of the items in the antecedent (http://www.resample.com/xl).

Interestingness measures are called fitness functions in Ykhlef (2011). The study in Ykhlef (2011) divides fitness functions into two types, basic and complex. Support and confidence are considered to be basic measures, whereas certain other fitness functions are derived from information theory and are considered to be complex fitness functions.

Many algorithms can be used to discover association rules from data to extract useful patterns. The Apriori algorithm is one of the most widely used and famous techniques for finding association rules (Agrawal et al., 1993; Agrawal, 1994). The Apriori algorithm requires two thresholds of minconfidence (representing minimum confidence) and minsupport (representing minimum support). These two thresholds determine the degree of association that must hold before a rule will be mined. The algorithm operates in two phases. In the first phase, all of the item sets with minimum support (frequent item sets) are generated. The second phase of the algorithm generates rules from the set of all frequent item sets.

1.2. Rule interestingness measures

Past research in data mining has shown that the interestingness of a rule can be measured using objective measures and subjective measures. Objective measures involve analyzing the rule's structure, predictive performance, and statistical significance. In association to rule mining, such measures include support and confidence (Liu et al., 2000). However, it is noted in Piatetsky-Shapiro and Matheus (1991) that such objective measures are insufficient for determining the interestingness of a discovered rule. Indeed, subjective measures are needed.

There are two main subjective interestingness measures, namely unexpectedness (Liu and Hsu, 1996; Silberschatz and Tuzhilin, 1996) and actionability (Piatetsky-Shapiro and Matheus, 1991; Silberschatz and Tuzhilin, 1996).

- Unexpectedness: Rules are interesting if they are unknown to the user or contradict the user's existing knowledge (or expectations).
- Actionability: Rules are interesting if the user can do something with them to his/her advantage.

In this research, we focus only on unexpectedness.

1.3. Ontology

The term ontology has been widely used in recent years in the field of Artificial Intelligence and computer and information science, especially in domains such as cooperative information systems, intelligent information integration, information retrieval and extraction, knowledge representation, and database management systems (Guarino, 1998). Although there is no universal consensus on the definition of an ontology, it is generally accepted that ontology is a specification of conceptualization (Leacock and Chodorow, 1998). The prior knowledge of a domain or a process in the field of data mining can help to select the appropriate information (preprocessing), decrease the space of the hypotheses (processing), represent results in a more comprehensible way and improve processing (or post-processing) (Farzanyar et al., 2006). Ontologies express the domain knowledge, which includes the semantic links between domain individuals that are described as relations of inter-concepts or roles (Gruber, 1995). Ontologies are usually constructed by domain experts, which results in a consensual and shared knowledge and allows domain knowledge to be captured in an explicit and formal way such that it can be shared among human and computer systems. The study in Corcho et al. (2003) distinguishes between lightweight ontologies, which include concepts, concept taxonomies, relationships between concepts and properties that describe concepts, from heavyweight ontologies, which add axioms and constraints to lightweight ontologies.

Ontologies are encoded in a formal language called an ontology language. There are several languages that are used for expressing ontologies, such as KIF, Loom, OCML, FLogic, RDF, RDF (S), and OWL (Web Ontology Language).

1.4. Fuzzy ontology

Knowledge about a system contains ambiguity and vagueness. Fuzzy sets were introduced by Zadeh in Zadeh, (1965) as a mathematical tool to solve the problem of vagueness. It is convenient to represent knowledge using fuzzy sets and fuzzy relations. The fuzzy ontology has been introduced to represent fuzzy concepts and relationships where concepts are related to others in the ontology with a degree of membership μ ($0 \leqslant \mu \leqslant 1$) assigned to the relationship. The fuzzy ontology is a hierarchical relationship between concepts within a domain, which can be viewed as a graph. It is developed based on the ontology graph and fuzzy logic. Fuzzy ontology captures richer semantics than traditional domain knowledge representations by allowing partial belonging of one item to another.

Fig. 1 shows a concept hierarchy of food items based on the taxonomy presented in Chen et al. (2000), where 'Tomato' can be regarded as being both 'Fruit' and 'Vegetable', but to different degrees.

Our approach uses the fuzzy membership degree in "IS-A" relationships between concepts.

1.5. Conceptual distance

Two main categories of algorithms for computing the semantic distance between terms organized in a hierarchical structure have been proposed in the literature (Jiang and Conrath., 1997): distance-based approaches and information content-based approaches. The general idea behind the distance-based algorithms (Leacock and Chodorow, 1998; Rada et al., 1989; Wu and Palmer, 1994) is to find the shortest path between two concepts in terms of the number of edges. The shorter the path from one node to the other, the more similar they are. The problem with this approach is that it relies on the notion that edges in a taxonomy represent uniform distances (i.e., it assumes that all conceptual links are of equal weight). Information content-based approaches (Rada et al., 1989; Jiang and Conrath., 1997) are inspired by the perception that pairs of concepts that share many common contexts are semantically related. The more information that two concepts share in common, the more similar they are.

The problem of the ontology distance is that it is highly dependent on the construction of the ontology. The measure is, therefore, highly dependent on oftentimes subjective ontology engineering decisions. To address this problem, we are associating a weight to any concept in the ontology that represents the degree of importance of this concept in the ontology along with the strength of any relation between the concepts. In an IS-A semantic network, the simplest form of determining the distance between two concept nodes, A and B, is the shortest path that links A and B, i.e., the minimum number of edges that separate A and B (Rada et al., 1989) or the sum of the weights of the arcs along the shortest path between A and B (Richardson et al., 1995). In the hierarchy of Fig. 1, the edge distances are:

$$\mathrm{Dist}(\mathrm{Apple}, \ \mathrm{Kiwi}) = 2 \qquad \mathrm{Dist}(\mathrm{Carrots}, \ \mathrm{Pepper}) = 2$$
$$\mathrm{Dist}(\mathrm{Apple}, \ \mathrm{Meat}) = 4 \qquad \mathrm{Dist}(\mathrm{Fruit}, \ \mathrm{Red\ Meat}) = 4$$

2. Related studies

The unexpectedness of patterns has been studied in Liu and Hsu (1996)Silberschatz and Tuzhilin, 1996(Liu et al., 1997)Padmanabhan and A., 1998(Padmanabhan, 1999)

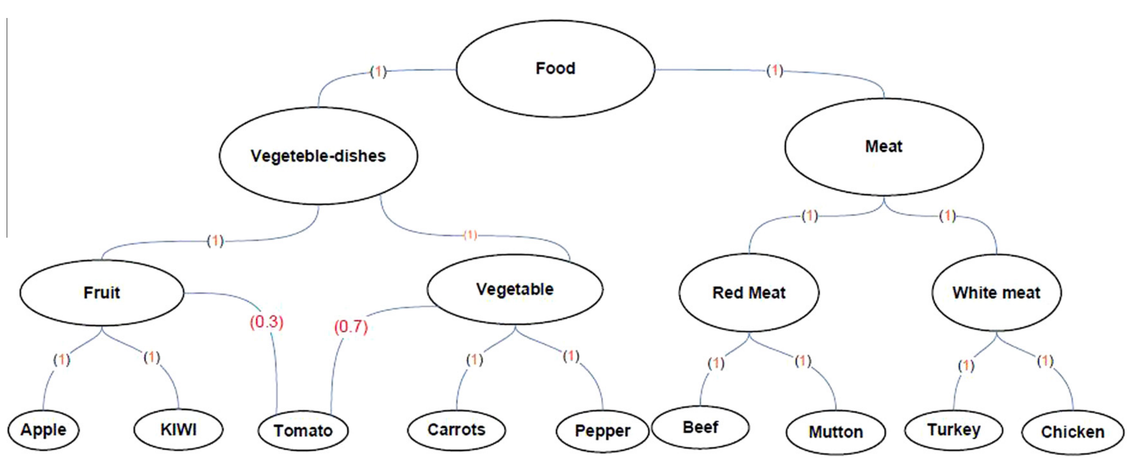

Figure 1 Fuzzy hierarchy example.

(Silberschatz and Tuzhilin, 1995 and defined in comparison with user beliefs. A rule is considered to be interesting if it affects the levels of conviction of the user. The unexpectedness is defined in probabilistic terms in Silberschatz and Tuzhilin (1996)Silberschatz and Tuzhilin, 1995, while in Liu and Hsu (1996), it is defined as a distance, and it is based on a syntactic comparison between a rule and a conviction. Similarity and distance is defined syntactically based on the structure of the rules and convictions. A rule and a conviction are distant if the consequence of the rule and conviction is similar but the antecedents are distant, or vice versa. In Padmanabhan and Tuzhilin, 2006, the focus is on discovering minimal unexpected patterns rather than using any of the post-processing approaches, such as filtering, to determine the minimal unexpected patterns from the set of all of the discovered patterns. In Padmanabhan and **A.Tuzhilin, 1997, unexpectedness is defined from the point of view of a logical contradiction of a rule and conviction; the pattern that contradicts prior knowledge is unexpected. This concept is based on the contradiction of the consequence of the rule and the consequence of belief. Given a rule $A \rightarrow B$ and a belief $X \rightarrow Y$, if B AND Y is False while A AND X is true for broad group of data, then the rule is unexpected. In Liu et al., 1999, the subjective interestingness (unexpectedness) of a discovered pattern is characterized by asking the user to specify a set of patterns according to his/her previous knowledge or intuitive feelings. This specified set of patterns is then used by a fuzzy matching algorithm to match and rank the discovered patterns. Reference (Klemettinen et al., 1994) proposes a template-based approach in which the user specifies interesting and uninteresting association rules using templates. A template describes a set of rules in terms of items that occurred in the conditional and the consequent parts. The system then retrieves the matching rules from the set of discovered rules. The studies in Srikant et al. (1997)T Ng et al., 1998 propose an association rule mining algorithm that can take item constraints specified by the user in the rule mining process in such a way that only those rules that satisfy the constraints are generated. References (Sahar, 1999)Sahar et al., 2001(Sahar'', 2002) have taken a different approach to the discovery of interesting patterns by eliminating non-interesting association rules. Rather than getting the users to define their entire knowledge of a domain, they are asked to identify several non-interesting rules that were generated by the Apriori algorithm. The study in Sahar, (2002) uses a genetic algorithm to dynamically maintain and search populations of rule sets for the most interesting rules rather than act as a post-processor. The rules identified by the genetic algorithm compared favorably with the rules selected by the domain expert (McGarry, 2005). To find subjectively interesting rules, most existing approaches ask the user to explicitly specify what types of rules are interesting and uninteresting, then generate or retrieve those matching rules. This research on the unexpectedness makes a syntactic or semantic comparison between a rule and a belief.

3. Contributions

Past research in data mining has shown that the interestingness of a rule can be measured using objective measures and subjective measures. Lightweight ontologies include concepts, concept taxonomies, relationships between concepts and properties that describe concepts. We define the conceptual distance of a rule as the distance between its antecedent and consequent relative to a hierarchy of concepts. By ranking rules according to their conceptual distance, highly related concepts are grouped together in the hierarchy. The farther away the concepts are, the less related they are to each other. For concepts that take part in the definition of a rule, the less related the concepts are to each other, the more the rule is surprising and, therefore, interesting. With such a ranking method, a user can check fewer rules on the top of the list to extract the most pertinent ones

The basic idea of our technique comprises the generation of association rules using any algorithm of rule generation (the 'A-priori' Algorithm, for example) and adjusting the objective measures, such as support and confidence, to the user's needs. The output, i.e., the association rules that result from this process becomes the input of our algorithm along with the domain ontology. Our technique analyzes the discovered rules and computes their conceptual distance. The higher the distance is, the more this rule becomes interesting. Our definition of unexpectedness is based on the structure of the background knowledge (hierarchy) that underlies the terms (vocabulary) of the rule, which is the conceptual distance between the head and the body of the rule. We are taking a different approach from all of the preceding work. The preceding work is a filtering process where a belief is expressed as rules that the user must enter, as for a query system. We are proposing a ranking process, and the knowledge is not expressed as rules; instead, the knowledge is expressed as a hierarchy of ontology concepts. Our approach is giving intelligence and autonomy to the computer to rank the more interesting rules on the top of the list based on the background knowledge represented by the domain ontology. Ontologies enable knowledge sharing. Sharing vastly increases the potential for knowledge reuse and therefore allows our approach to obtain free knowledge solely from using domain ontologies that are already available, such as ''ONTODerm'' for dermatology, ''BIO-ONT'' for biomedicine, and ''AGROVOC'' for food.

4. Method presentation

Data mining is the process of discovering patterns in data. Data mining methods have a drawback in that they generate a very large number of rules that are not of interest to the user. The use of objective measures of interestingness, such as confidence and support, is a step toward interestingness. Objective measures of interestingness are data driven; they measure the statistical strength of the rule and do not exploit the domain knowledge and intuition of the decision maker. In addition to objective measures, our approach exploits domain knowledge that is represented by Fuzzy ontology organized as a DAG hierarchy. The nodes of the hierarchy represent the rules of the vocabulary. For a rule such as (x AND $y \rightarrow z$), x, y and z are nodes in the hierarchy. The conceptual distance between the antecedent (x AND y) and the consequent (z) of a rule is a measure of interestingness. The larger the distance is, the more the rule is unexpected and, therefore, interesting. Based on this measure, a ranking algorithm helps to select those rules that are of interest to the user.

The basic idea of our technique comprises the generation of association rules using any algorithm for rule generation (the

'A-priori' Algorithm, for example) and adjusting the objective measures, such as support and confidence, to the user's needs. The output, i.e., the association rules that result from this process becomes the input of our algorithm along with the domain ontology. Our technique analyzes the discovered rules and computes their conceptual distance. The larger the distance is, the more this rule is interesting. A concept in the rule that has no equivalent in the domain ontology (unknown concepts) can be either a noisy concept or a new concept. In such a case, the user can update the ontology based on new concepts, to produce ontology evolution (Fig. 2).

4.1. Concept semantic distance

In our approach, we are making a distinction between the weight that is associated with a concept (in Fig. 1, the weight is 1 for all of the concepts) and the strength of the relation between a concept child and its parent (the membership degree μ of the concept's child to its parent). In Fig. 1, $\mu = 1$ for all of the concepts except for 'Tomato'.

The semantic distance between the two concepts A and B is the sum of the weights of the arcs along the shortest path between A and B (Richardson et al., 1995). To calculate the weight for fuzzy relations, an extension to the weighting function is required.

The weighting function for the crisp hierarchy relations is f (μ, ω): {0, 1} xIR → IR where:

$$f(\omega) = \begin{cases} \omega & \mu = 1(\text{if} \quad A, B \text{ Connected}) \\ 0 & \mu = 0 \ (\text{if} \quad A, B \text{ Disconnected}) \end{cases} \tag{1}$$

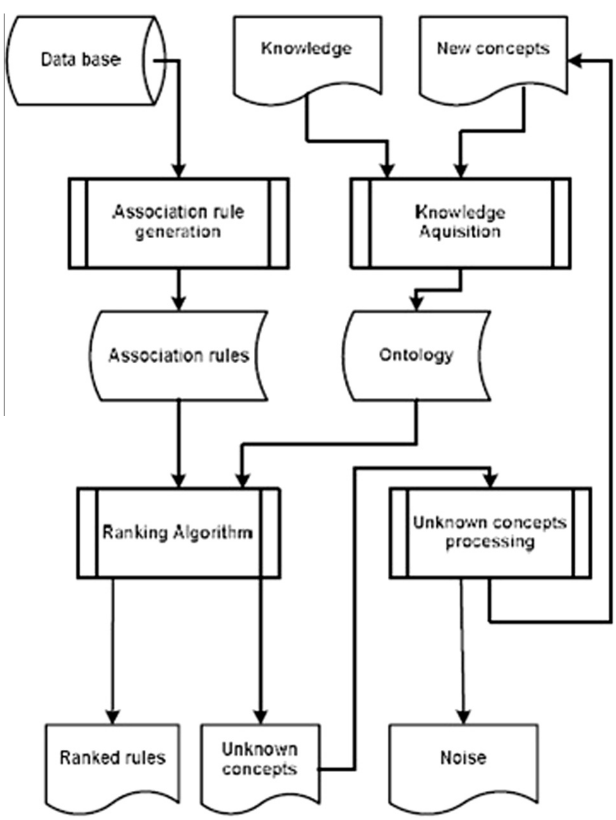

Figure 2 Ontology-driven rule ranking system.

$\mu \in \{0, 1\}$ is the membership degree of a concept to its parent, and ω is the weight associated with this concept in the hierarchy.

The function based on the Boolean variable ($\mu \in$ {0, 1}) in (1) is extended to a weighting function based on a continuous variable μ ∊ (Fayyad et al., 1996) in (2), according to the weight ω associated with the child concept and its degree of membership μ.

$$f(\mu, \omega) : [0, 1] \times \ \text{IR} \rightarrow \ \text{IR} \quad f(\mu, \omega) = \begin{cases} \omega + (1 - \mu) * \omega & \mu \neq 0 \\ 0 & \mu = 0 \end{cases} \tag{2}$$

where μ ($0 \leqslant \mu \leqslant 1$) is the membership degree.

For a degree membership $\mu = 1$, our extended weighting function is equal to the Boolean weighting function.

The weight of a concept (ω) is based on the density of the hierarchy for this concept and its depth in the hierarchy, to address the widely recognized problem of "edge-counting" (uniformity in the link distances of the taxonomy).

To compute the shortest path between two nodes, we use Dijkstra's algorithm (Dijkstra', 0000).

4.2. Rule conceptual distance

To compute the distance between groups of concepts, for a given rule R: $X \rightarrow Y$, where $X = X_1 \wedge \ldots \wedge X_k$, $Y = Y_1 \wedge \ldots \wedge Y_m$, we use the Hausdorff distance.

$$\text{Distance}(x, y) = \max(h(x, y), \max(y, x)) \text{ where } h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\| \tag{3}$$

The function $h(X, Y)$ is called the directed Hausdorff 'distance' from X to Y (this function is not symmetric and thus is not a true distance). This function identifies the point X_iX that is farthest from any point of Y and measures the distance from X_i to its nearest neighbor in Y. The Hausdorff distance, $H(X, Y)$, measures the degree of mismatch between two sets because it reflects the distance of the point of X that is farthest from any point of Y, and vice versa (Huttenlocher et al., 1993).

This expression measures the conceptual distance between groups $X = X_1 \wedge \ldots \wedge X_k$ and $Y = Y_1 \wedge \ldots \wedge Y_m$ of the concepts that contain the k X_i and m atomic Y_j concepts, respectively.

4.3. Rule ranking algorithm

In this section, we introduce an algorithm to rank the rules according to their conceptual distance based on a fuzzy ontology that represents the background knowledge. The rules that we consider are in the form of "body → head", where "body" and "head" are conjunctions of concepts in the vocabulary of the ontology. We assume that other techniques carry out the task of pattern discovery and eliminate the patterns that do not satisfy the objective criteria.

With such a ranking, a user can check only the patterns that are on the top of the list to confirm the rules that are the most pertinent. The algorithm will use the procedure to compute the weight of the edge based on the membership degree, and it will apply the proposed weighting function.

4.3.1. Summary of the Algorithm

Input

Rule set
Ontology

Processing

Build inter-concept semantic distance matrix.
For each rule, compute the rule conceptual distance
Sort the rule set

Output

Ranked rule set.

4.3.2. Pseudo code

Procedure weight (Shortest Path (X_i, X_j))

/* this procedure (which is called by the algorithm) calculates the weight of each edge in the path based on the membership degree and the original weight of the child node and sums up the weights of the path going from X_i to X_j. */

```
begin
  for Xk = Xi to Xj step 2
  begin
    //Compute the weight of the edge.
    w(Xk,Xk+1) = (ωXk + ωXk* (1−μ));
    //sum up the weight
    Total_Weight = Total_Weight + w(Xk,Xk+1);
  end
  return (Total_Weight);
end
```

Algorithm

```
ND: Number of nodes
R: Set of rules R = {Ri/Ri = body → head} where i ∈ [1,N]
N: number of rules
D: Maximum depth of the hierarchy
Xi, Yj: Atomic Concepts; i ∈ [1,k]; j ∈ [1,m]
Body = X1∧...∧Xk
Head = Y1∧...∧Ym
//for all of the nodes in the graph, calculate the conceptual distance
for i = 1 to ND
for j = 1 to ND
begin
  // ShortestPath(Xi,Xj) shortest path between Xi and Xj
  //Make a call to the weight(ShortestPath(Xi,Xj) above.
  Distance(Xi,Xj)= weight(ShortestPath(Xi,Xj));
end
//rule conceptual distance computation
for i = 1 to N
  Distance(Ri) = (Distance(X1∧...∧Xk; Y1∧...∧Ym));
Sort Distance(Ri) descending;
```

5. Example

For a set of association rules R = {(a) Apple → Kiwi; (b) Apple → Carrots; (c) Pepper, Carrots → Turkey, Chicken; (d) Kiwi → Tomato; (e) Tomato → Pepper; (f) Tomato, Pepper → Turkey, Chicken}, we apply our method to the hierarchy of Fig. 1. Fig. 3 represents the indexed table of concepts. In Fig. 4, every cell in the semantic distance inter-concepts of the ontology represents the distance between the concept in the row and the corresponding concept in the column. The results are shown in Fig. 5.

Inter-concept semantic distance Computation

The number of nodes in (Fig. 1) is ND = 16. For $\omega = 1$, the weighting function

$$f(\mu,\omega) : [0,1] \times \mathrm{IR} \rightarrow I \; f(\mu,1) = \begin{cases} 1 + (1-\mu) & \mu \neq 0 \\ 0 & \mu = 0 \end{cases} \quad (4)$$

Let $w\ (i, j)$ be the weight between node i and node j; then,

$$\mathrm{w(tomato, fruit)} = 1 + (1 - \mu(\mathrm{tomato, fruit})) = 1 + (1\text{–}0.3) = 1.7.$$

$$\mathrm{w(tomato, Vegetable)} = 1 + (1 - \mu(\mathrm{tomato, Vegetable})) = 1 + (1\text{–}0.7) = 1.3$$

The shortest path between any two concepts is the conceptual distance between two of them.

Let Dist (i, j) be the semantic distance between concept i and concept j; then,

$$\mathrm{Dist(Apple, Kiwi)} = \mathrm{w(Apple, Fruit)} + \mathrm{w(Fruit, Kiwi)} = 2.$$

$$\mathrm{Dist(Apple, Tomato)} = \mathrm{w(Apple, Fruit)} + \mathrm{w(Fruit, Tomato)} = 2.7$$

Rule conceptual distance computation

For the set of rules R = {(a), (b), (c), (e), and (f)} where:

(a) Apple → Kiwi;
(b) Apple → Carrots;
(c) Pepper, Carrots → Turkey, Chicken;
(d) Kiwi → Tomato;
(e) Tomato → Pepper;
(f) Tomato, Pepper → Turkey, Chicken

By applying the Hausdorff distance, the rule distances are:

(a) Dist (Apple, Kiwi) = 2
(b) Dist (Apple, Carrots) = 4
(c) Dist (Pepper ∧ Carrots, Turkey ∧ Chicken) = 6

0	Apple	8	Chicken
1	Kiwi	9	Fruit
2	Tomato	10	Vegetable
3	Carrots	11	Red Meat
4	Pepper	12	White Meat
5	Beef	13	Vegetable-dishes
6	Mutton	14	Meat
7	Turkey	15	Food

Figure 3 Concepts' index table.

0.0	2.0	2.7	4.0	4.0	6.0	6.0	6.0	6.0	1.0	3.0	5.0	5.0	2.0	4.0	3.0
2.0	0.0	2.7	4.0	4.0	6.0	6.0	6.0	6.0	1.0	3.0	5.0	5.0	2.0	4.0	3.0
2.7	2.7	0.0	2.3	2.3	6.3	6.3	6.3	6.3	1.7	1.3	5.3	5.3	2.3	4.3	3.3
4.0	4.0	2.3	0.0	2.0	6.0	6.0	6.0	6.0	3.0	1.0	5.0	5.0	2.0	4.0	3.0
4.0	4.0	2.3	2.0	0.0	6.0	6.0	6.0	6.0	3.0	1.0	5.0	5.0	2.0	4.0	3.0
6.0	6.0	6.3	6.0	6.0	0.0	2.0	4.0	4.0	5.0	5.0	1.0	3.0	4.0	2.0	3.0
6.0	6.0	6.3	6.0	6.0	2.0	0.0	4.0	4.0	5.0	5.0	1.0	3.0	4.0	2.0	3.0
6.0	6.0	6.3	6.0	6.0	4.0	4.0	0.0	2.0	5.0	5.0	3.0	1.0	4.0	2.0	3.0
6.0	6.0	6.3	6.0	6.0	4.0	4.0	2.0	0.0	5.0	5.0	3.0	1.0	4.0	2.0	3.0
1.0	1.0	1.7	3.0	3.0	5.0	5.0	5.0	5.0	0.0	2.0	4.0	4.0	1.0	3.0	2.0
3.0	3.0	1.3	1.0	1.0	5.0	5.0	5.0	5.0	2.0	0.0	4.0	4.0	1.0	3.0	2.0
5.0	5.0	5.3	5.0	5.0	1.0	1.0	3.0	3.0	4.0	4.0	0.0	2.0	3.0	1.0	2.0
5.0	5.0	5.3	5.0	5.0	3.0	3.0	1.0	1.0	4.0	4.0	2.0	0.0	3.0	1.0	2.0
2.0	2.0	2.3	2.0	2.0	4.0	4.0	4.0	4.0	1.0	1.0	3.0	3.0	0.0	2.0	1.0
4.0	4.0	4.3	4.0	4.0	2.0	2.0	2.0	2.0	3.0	3.0	1.0	1.0	2.0	0.0	1.0
3.0	3.0	3.3	3.0	3.0	3.0	3.0	3.0	3.0	2.0	2.0	2.0	2.0	1.0	1.0	0.0

Figure 4 Inter-concepts' semantic distance.

[Distance=6.3]	[Turkey, Chicken] <- [Pepper, Tomato]
[Distance=6]	[Turkey, Chicken] <- [Pepper, Carrots]
[Distance=4]	[Carrots] <- [Apple]
[Distance=2.7]	[Tomato] <- [Kiwi]
[Distance=2.3]	[Pepper] <- [Tomato]
[Distance=2]	[Kiwi] <- [Apple]

Figure 5 Example rules ranking results.

(d) Dist (Kiwi, Tomato) = 2.7
(e) Dist (Tomato, Pepper) = 2.3
(f) Dist (Tomato $\wedge$ Pepper, Turkey $\wedge$ Chicken) = 6.3

The order of rules would be (f), (c), (b), (d), (e), (a) based on taking the conceptual distance in a descending order, as shown in Fig. 5. From the perspective of the decision system, the rules (f) and (c) belong to a higher level (food) than the rules (b) and (d), which belong to a lower level (vegetable-dishes). The rules (e) and (a) belong to a lower level (vegetable) and (fruit), respectively. The more that we move up in the hierarchy, the more the decision is important, and the vision of the decision maker is broader and therefore the discovered rule is more interesting. Rules (f) and (c) are the crossing result of the domains (vegetables-dishes, meat), which are farther than the domains (vegetables, fruits) of the rule (b) and (d). The rules (e) and (a) concern only the domain (vegetable) and (fruit), respectively. Therefore, they are less interesting. Note that rule (d) is more surprising than rule (e), even though tomato is a fruit and vegetable with different degrees. Because a tomato is closer to a vegetable than a fruit, rule (d) is more interesting than rule (e).

6. Experiments

The experiments were performed using a census income database (in, 0000) and an implementation of our algorithm. To generate the association rules, we used the implementation of the 'Apriori' algorithm (Christian, 0000) with a minimum support value equal to 0.2 and a minimum confidence value equal to 0.2. The number of generated rule sets is 2225. To perform the experiments, we created a taxonomy of 81 weighted concepts (Fig. 6) based on the data set that we are studying, and we defined two fuzzy concepts, 'Low_Level' and 'High_Level', for education as fuzzy sets.

The membership functions to the fuzzy sets are:

$\mu_{\text{Hight}}(\text{Level\_Rank}) = \text{Level\_Rank}/15$
$\mu_{\text{Low}}(\text{Level\_Rank}) = 1 - \mu_{\text{Hight}}(\text{Level\_Rank}) = \text{Level\_Rank}/15$, where Level_Rank is a sequential number that ranges from 0 to 15, where 0 represents 'Preschool' and 15 represents 'Doctorate'.
Level_Rank = {(Preschool = 0), (1st-4th = 1), (5th-6th = 2), (7th-8th = 3), (9th = 4), (10th = 5), (11th = 6), (12th = 7), (HS-grad = 8), (Some-college = 9), (Assoc-acdm = 10), (Assoc-voc = 11), (Bachelors = 12), (Prof-school = 13), (Masters = 14), (Doctorate = 15)}

Formally, High_Level and Low_Level fuzzy sets can be defined as:

Level 0 Census-Income	Level 1	Level 2	Atomic concepts	Weight	μ
	Work			1	1.00
		WorkClass		1	1.00
			Private	1	1.00
			Self-emp-not-inc	1	1.00
			:	1	1.00
			Never-worked	1	1.00
		Occupation		1	1.00
			Tech-Support	2	1.00
			Craft-Repair	2	1.00
			:	2	1.00
			Armed-Forces	2	1.00
		SalaryClass		1	1.00
			>50K	1	1.00
			<=50K	1	1.00
	EDUCATION			1	1.00
		HighLevel		1	1.00
			Preschool	7	0.00
			1st-4th	7	0.07
			:	7	
			Masters	7	0.93
			Doctorate	7	1.00
		LowLevel		1	1.00
			Preschool	7	1.00
			1st-4th	7	0.93
			:	7	
			Masters	7	0.07
			Doctorate	7	0.00
		EducationNum		1	1.00
			<=9	1	1.00
			9< Education<=13	1	1.00
			13<num<=15	1	1.00
			>15	1	1.00
	Personal			1	1.00
		Race		1	1.00
			White	2	1.00
			Black	2	1.00
			Asian-Pac-Islander	2	1.00
			Amer-Indian-Eskimo	2	1.00
			Other	2	1.00
		Sex		1	1.00
			Female	1	1.00
			Male	2	1.00
		Native-Country		1	1.00
			Europe	1	1.00
			Asia	1	1.00
			SouthAmerica	1	1.00
			United-States	1	1.00
			Canada	1	1.00
		Age		1	1.00
			age <=20	1	1.00
			20<age <=30	1	1.00
			:	1	
			>50	1	1.00
		Marital-Status		1	1.00
			Married-civ-spouse	2	1.00
			:	2	1.00
			Married-AF-spouse.	2	1.00

Figure 6 Census-income ontology.

High_Level = {Preschool/0.00, 1st-4th/0.07,
5th-6th/0.13, 7th-8th/0.20,
9th/0.27, 10th/0.33,
11th/0.40, 12th/0.4
HS-grad /0.53, Some-college/0.60,
Assoc-acdm /0.67, Assoc-voc /0.73,
Bachelors /0.80, Prof-school /0.87,
Masters /0.93, Doctorate/1.00}

[Distance=15.29]	[Craft-repair] <- [HS-grad]
[Distance=15.29]	[HS-grad] <- [Never-married]
[Distance=14.8]	[Never-married] <- [Some-college]
[Distance=14.8]	[Some-college] <- [Never-married]
[Distance=14.29]	[HS-grad] <- [age>50]
[Distance=14.29]	[age>50] <- [HS-grad]
[Distance=13.8]	[40<age<=50] <- [Some-college]
[Distance=13.8]	[Some-college] <- [40<age<=50]
[Distance=13.4]	[Bachelors] <- [Never-married]
[Distance=13.4]	[Bachelors] <- [>50K, Married-civ-spouse]
[Distance=13.29]	[EducationNum<=9] <- [HS-grad]
[Distance=13.29]	[HS-grad] <- [EducationNum<=9]
[Distance=12.8]	[9<EducationNum<=13] <- [Some-college]
[Distance=12.8]	[Some-college] <- [9<EducationNum<=13]
[Distance=12.4]	[Bachelors] <- [>50K]
[Distance=12.4]	[Bachelors] <- [>50K, Male]
[Distance=8]	[Adm-clerical] <- [Female]
[Distance=8]	[Adm-clerical] <- [Female, Private]
[Distance=7]	[Married-civ-spouse] <- [>50K]
[Distance=7]	[>50K] <- [Married-civ-spouse]
[Distance=6]	[Craft-repair] <- [EducationNum<=9]
[Distance=6]	[Divorced] <- [Female]
[Distance=5]	[Exec-managerial] <- [>50K]
[Distance=5]	[Prof-specialty] <- [>50K]
[Distance=4]	[Male] <- [age>50]
[Distance=4]	[age>50] <- [Male]

Figure 7 Rules ranking results for the first experiment.

[Distance=6]	[Adm-clerical] <- [Female]
[Distance=6]	[>50K] <- [Male]
[Distance=5]	[Craft-repair] <- [EducationNum<=9]
[Distance=5]	[EducationNum<=9] <- [>50]
[Distance=4.47]	[Craft-repair] <- [HS-grad]
[Distance=4.47]	[HS-grad] <- [>50]
[Distance=4.4]	[40<age<=50] <- [Some-college]
[Distance=4.4]	[>50K] <- [Some-college]
[Distance=4.2]	[Bachelors] <- [>50K]
[Distance=4.2]	[Bachelors] <- [Never-married]
[Distance=4]	[Exec-managerial] <- [>50K]
[Distance=4]	[Prof-specialty] <- [>50K]
[Distance=3.47]	[EducationNum<=9] <- [HS-grad]
[Distance=3.4]	[9<EducationNum<=13] <- [Some-college]

Figure 8 Rules ranking results for the second experiment.

Low_Level = {Preschool/1.00, 1st-4th/0.93, 5th-6th/0.87, 7th-8th/0.80, 9th/0.73,10th/0.67, 11th/0.60, 12th/0.53, HS-grad /0.47, Some-college/0.40, Assoc-acdm /0.33, Assoc-voc /0.27, Bachelors /0.20, Prof-school /0.13, Masters /0.07, Doctorate/0.00}

We have conducted two experiments on our dataset. The first experiment was conducted with a weight equal to 7, for all of the atomic concepts related to education, and the second experiment was conducted with a weight equal to 1, for all atomic concepts.

6.1. First experiment

In this experiment, a test was conducted with a weight equal to 7 for all of the atomic concepts that were related to education and with different weights for the remaining concepts (Fig. 6). The user in this case is placing more emphasis on educational concepts by setting their weights to a high value or because they are truly important in the domain of study. The ranking rules algorithm chooses those with higher weights. The results are presented in Fig. 7.

6.2. Second experiment

In this experiment, the weight is set to 1 for all of the atomic concepts. All of the concepts are equally important for the user. The results are shown in Fig. 8.

6.3. Discussion

The first experiment was conducted with a weight that was set to 7 for all of the atomic concepts related to education. The user in this case is placing more emphasis on these concepts by setting their weights to a high value.

Subsume the concepts of 'HS-grade' and 'craft-repair' of rule (1) in Fig. 7 as 'census-income' (see Fig. 6). These concepts are less related to each other based on the ontology. The same reasoning applies to rule (2) in Fig. 7.

Subsume the concept of 'some-college' and 'Never-Married' of rule (3) in Fig. 7 as 'census-income' also; however, the membership degree of 'some-college' is 0.60 (Fig. 6). Note that the weight for both concepts, 'craft-repair' and 'Never-Married', is 2.

'HS-grade' has a membership degree of 0.53 and is less than the membership degree of the concept 'some-college', which is 0.60 (rule (3)); this value makes this degree relatively far from the expected and therefore more interesting.

The common subsumption for the last 2 rules of Fig. 7 is the concept 'Personal' (Fig. 6). These rules express the relation between the concepts 'sex' and 'age', and they are close to each other based on the ontology.

The second experiment was conducted with a weight that was set to 1 for all of the atomic concepts. The user in this case is considering all of the concepts to be equally important.

Subsume the concepts 'Adm-Clirical' and 'Female' of rule (1) in Fig. 8 as 'census-income' (Fig. 6). 'Adm-Clirical' is a child of the 'work' concept, whereas 'sex' is a parent of the concept 'Female'. These concepts are less related to each other based on the ontology. 'Adm-Clirical' has a membership degree of 1 and is the same as the 'Female' concept. The same reasoning is applied to the rules (2), (3) and (4) in Fig. 8.

Subsume the concepts 'Craft-repair' and 'HS-grade' of rule (5) as 'census-income' as well, but 'HS-Grade' has a membership degree of 0.53, which is less than the membership degree of the concepts within rules (1), (2), (3) and (4).

The common subsumption for the last 2 rules of Fig. 8 is the concept 'Education' (Fig. 6). These rules express the relation between the concepts 'EducationNum' and 'EducationLevel', which are close to each other based on the ontology.

A rule such as (1) or (2) in Fig. 7 is more interesting because it is giving us information between the 'Education' and 'Occupation' information and it involves a higher decision maker (strategic) than the last two rules, which concern 'sex' and 'age'. The same reasoning applies for Fig. 8. The more that we move up in the hierarchy, the more the decision is important and the vision of the decision maker is broader, strategic and important; therefore, the discovered rule is more interesting.

7. Conclusions and future work

In this paper, we proposed a new approach for ranking association rules according to their conceptual distance, which was defined on the basis of the ontological distance. The proposed ranking algorithm helps the user to identify interesting association rules, particularly expected and unexpected rules. This algorithm uses a fuzzy lightweight ontology to calculate the distance between the antecedent and the consequent of the rules on which the ranking is based. The larger the conceptual distance is, the more the rule represents a high degree of interest. We proposed an extension to the mapping weighting function based on the membership degree to compute the weight of the relations in the fuzzy ontology. In the future, we plan to integrate the concept proprieties in the conceptual distance computing and exploit other relation types of the ontology. We are planning to integrate a personal ontology that represents a user's views and interests into a domain ontology to develop ranking rules that are based on the integrated output.

Acknowledgments

The authors would like to thank the anonymous reviewers for their constructive comments. These comments have contributed to a vast improvement of this paper.

References

Agrawal, R., Srikant, R., 1994. Fast algorithms for mining association rules. Proceedings of the 20th International Conference Very Large Data Bases 1215, 487–499.

Agrawal, R., Imieliński, T., Swami, A., 1993. Mining association rules between sets of items in large databases'', ACM SIGMOD Record, vol. 22(2), pp. 207–216.

Agrawal, R., Imielinski, T., Swami, A., 1993. Database mining: A performance perspective. Knowl. Data Eng. IEEE Trans. 5 (6), 914–925.

Asa, J.A.M., Mangano, J.J., 1995. Selecting among rules induced from a hurricane database. J. Intelligent Inf. Sys. 4 (1), 39–52.

Census income. //ftp.ics.uci.edu/pub/machine-learning-databases/census-income/. Accessed ''05/25/2012''.

Chen, G., Wei, Q., Kerre, E.E., 2000. Fuzzy data mining: discovery of fuzzy generalized association rules. Studies Fuzziness Soft Comput. 53, 45–66.

Christian Borgelt's Web Pages. http://www.borgelt.net/software.html. Accessed 05/25/2012.

Corcho, O., Fernández-López, M., Gómez-Pérez, A., 2003. Methodologies, tools and languages for building ontologies. Where is their meeting point? Data Knowl. Eng. 46 (1), 41–64.

Dijkstra's algorithm. http://en.wikipedia.org/wiki/Dijkstra%27s_algorithm. Accessed 05/25/2012.

Farzanyar, Z., kangavari, M., Hashemi, S., 2006. A new algorithm for mining fuzzy association rules in the large databases based on ontology, Proceedings of the Sixth IEEE International Conference on Data Mining, pp. 65–69, .

Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., 1996. From data mining to knowledge discovery in databases. AI Mag. 17 (3), 37–54.

Gruber, T.R., 1995. Toward principles for the design of ontologies used for knowledge sharing. Int. J. Hum. Comput. Studies 43 (5), 907–928.

Guarino, N., 1998. Formal Ontology in Information Systems. Proceedings of the 1st International Conference of formal ontology in, information systems.

http://www.resample.com/xlminer/help/Assocrules/association-rules_intro.htm. Accessed 05/30/2012.

Huttenlocher, D.P., Klanderman, G.A., Rucklidge, W.J., 1993. Comparing images using the Hausdorff distance. IEEE Trans. Pattern Anal. Mach. Int. 15 (9), 850–863.

Jiang, J. J., Conrath, D. W. (1997). Semantic similarity based on corpus statistics and lexical taxonomy. arXiv, preprint cmp-lg/9709008.

Klemettinen, M., Mannila, H., Ronkainen, P., Toivonena, H., Verkamo, A.I., 1994. Finding interesting rules from large sets of discovered association rules, Proceedings of the third international conference on Information and, knowledge management, pp. 401–407.

Leacock, C., Chodorow, M., 1998. Combining local context and WordNet similarity for word sense identification, WordNet: An electronic lexical, database, 49(2), pp. 265–283.

Liu, B., Hsu, W. Post-analysis of learned rules Proceedings of the Thirteenth National Conference on Artificial Intelligence, pp. 828–834, 1996.

Liu, B., Hsu, W., Chen, S., 1997. Using general impressions to analyze discovered classification rules, Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining, pp. 31–36.

Liu, B., Hsu, W., Mun, L.F., Lee, H.Y., 1999. Finding interesting patterns using user expectations. IEEE Trans. Knowl. Data Eng. 11 (6), 817–832.

Liu, B., Hsu, W., Chen, S., Ma, Y., 2000. Analyzing the subjective interestingness of association rules. Intelligent Sys. Appl. IEEE 15 (5), 47–55.

McGarry, K., 2005. A survey of interestingness measures for knowledge discovery. Knowl. Eng. Rev. 20 (01), 39–61.

Padmanabhan, B., Tuzhilin, A., 1999. Unexpectedness as a measure of interestingness in knowledge discovery. Decision Support Sys. 27 (3), 303–318.

Padmanabhan, B., Tuzhilin, A., 1998. A belief-driven method for discovering unexpected patterns, Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining, pp. 94–100.

Padmanabhan, B., Tuzhilin, A., 1997. On the Discovery of Unexpected Rules in Data Mining Applications, Proceedings. of the Workshop on InformationTechnology and Systems, pp. 81–90.

Padmanabhan, B., Tuzhilin, A., 2006. On characterization and discovery of minimal unexpected patterns in rule discovery. IEEE Trans. Knowl. Data Eng. 18 (2), 202–216.

Piatetsky-Shapiro, G., 1996. Discovery analysis and presentation of strong rules. Knowl. Dis. Databases, 229–238.

Piatetsky-Shapiro, G., Matheus, C.J., 1991. The interestingness of deviations, Proceedings of the AAAI-94 workshop on Knowledge Discovery in Databases vol. 1, pp. 25–36.

Rada, R., Mili, H., Bicknell, E., Blettner, M., 1989. Development and application of a metric on semantic nets. IEEE Trans. Syst. Man Cybernetics 19 (1), 17–30.

Richardson, R., Smeaton, A.F., alan, F., 1995. Using WordNet in a knowledge-based approach to information retrieval", Technical Report CA-0395, School of Computer Applications, Dublin City University.

Sahar, S., 1999. Interestingness via what is not interesting", Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 332–336.

Sahar, S. 2001. Interestingness preprocessing. In Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on (pp. 489–496). IEEE.

Sahar, S., 2002. On incorporating subjective interestingness into the mining process, In Data Mining, 2002. ICDM 2003. Proceedings. Of the 2002 IEEE International Conference on Data Mining, pp. 681–684.

Silberschatz, A., Tuzhilin, A., 1995. On subjective measures of interestingness in knowledge discovery, Proceedings of th first International Conference on Knowledge Discovery and Data Mining, pp. 275–281.

Silberschatz, A., Tuzhilin, A., 1996. What makes patterns interesting in knowledge discovery systems. Knowl. Data Eng. IEEE Trans. 8 (6), 970–974.

Srikant, R., Vu, Q., Agrawal, R., 1997. Mining association rules with item constraints. In *KDD* 97, 67–73.

Ng, R.T., Lakshmanan, L.V., Han, J., Pang, A., 1998. Exploratory mining and pruning optimizations of constrained associations rules, ACM SIGMOD Record, Vol. 27(2), pp. 13–24.

Uthurusamy, R., Fayyad, U., Spangler, S., 1991. Learning useful rules from inconclusive data, Knowledge Discovery in Databases, Ed. G. Piatetsky-Shapiro, AAAI Press, pp. 141–15.

Widyantoro, D.H., Yen, J., 2001. A fuzzy ontology-based abstract search engine and its user studies. In Fuzzy Systems, Proceedings of the 10th IEEE International Conference on fuzzy systems, vol. 3, pp. 1291–1294.

Wu, Z., Palmer, M., 1994. Verbs semantics and lexical selection, Proceedings of the 32nd annual meeting on Association for Computational Linguistic, pp. 133–138.

Ykhlef, M., 2011. A quantum swarm evolutionary algorithm for mining association rules in large databases. J. King Saud University-Comput. Inf. Sci. 23 (1), 1–6.

Zadeh, L.A., Fuzzy sets, 1965. Information and control, vol. 8(3), 338–353.

Zimmermann, H.J., Fuzzy set theory-and its applications, Springer, 2001.