Fuzzy Ontology-Driven Method for Ranking Unexpected Rules

MOHAMED SAID HAMANI
Mohamed Boudiaf-M'sila University, Algeria
saidhamani@hotmail.com

RAMDANE MAAMRI
Mentouri-Constantine University, Algeria
rmaamri@yahoo.fr

ABSTRACT

Several rule discovery algorithms have the disadvantage to discover too much patterns sometimes obvious, useless or not very interesting to the user. The evaluation of patterns is based more on objective measures and is not based on prior background domain knowledge. Knowledge about the system contains ambiguity and vagueness. Fuzzy ontologies are able to deal with fuzzy knowledge representing knowledge using fuzzy sets and fuzzy relations. In this paper we propose a new approach for patterns ranking according to their unexpectedness and based on a prior background knowledge represented by domain fuzzy ontology organized as DAG (Directed Acyclic Graph) hierarchy.

Key words: data mining, fuzzy ontology, unexpectedness, association rules, domain knowledge, subjective measures, semantic distance.

1. Introduction

Knowledge discovery in databases (data mining) has been defined in [4] as the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns from data. Association rule algorithms [1] are rule-discovery methods that discover patterns in the form of IF-THEN rules. It was noticed that most algorithm of data mining generates a large number of rules who are valid but obvious or not very interesting to the user [19, 18, 22, and 10]. The presence of the huge number of rules makes it difficult for the user to identify those that are of interest. To address this issue most approaches on knowledge discovery use objective measures of interestingness, such as confidence and support [1], for the evaluation of the discovered rules. These objective measures capture the statistical strength of a pattern. The interestingness of a rule is essentially subjective [19, 22, 10, and 8]. Subjective measures of interestingness, such as unexpectedness [13, 23, and 2], 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 useful, they are insufficient in the determination of the interestingness of rules. One way to approach this problem is by focusing on discovering unexpected patterns [21, 22, 10, 11, 15 and 16] where unexpectedness of discovered patterns is usually defined relative to a system of prior expectations. In this paper we define unexpectedness based on the semantic distance of the rule vocabulary and relative to a prior knowledge represented by ontology. Ontology represents knowledge with the relationships between concepts. It is organized as a DAG (Directed Acyclic Graph) hierarchy. Ontologies allow domain knowledge to be captured in an explicit and formal way such that it can be shared among human and computer systems. Knowledge about a system contains ambiguity and vagueness. Fuzzy ontologies are able to deal with fuzzy knowledge [33] where concepts are related to each other in the ontology, with a degree of membership μ ($0 \leq \mu \leq 1$).We propose a new approach for ranking the most interesting rules according to conceptual distance (distance between the antecedent and the consequent of the rule) relative to the hierarchy. Highly related concepts are grouped together in the hierarchy. The more concepts are far away, the less are related to each other. The least concepts are related to each other and take part of the definition of a rule the more surprising the rule is and therefore interesting. With such ranking, a user can check fewer rules on the top of the list to extract the most pertinent ones.

2. Method Presentation

Data-mining is the process of discovering patterns in data. Data-mining methods have the drawbacks to 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 domain knowledge and intuition of the decision maker. Beside objective measures, our approach exploit domain knowledge represented by Fuzzy ontology organized as DAG hierarchy. The nodes of the hierarchy represent the rules vocabulary. For a rule like (x AND y→z) x, y and z are nodes in the hierarchy. The semantic distance between the Antecedent (x AND y) and the consequent (z) of the rule is a measure of interestingness. The more the distance is high, the more the rule is unexpected and therefore interesting. Based on this measure a ranking algorithm helps in selecting those rules of interest to the user.

2.1. Crisp Ontology

Although there is not a universal consensus on the definition of ontology, it is generally accepted that ontology is a specification of conceptualization [9]. Ontology can take the simple form of a taxonomy (i.e., knowledge encoded in a minimal hierarchical structure) or as a vocabulary with standardized machine interpretable terminology. The prior knowledge of domain or a process in the field of data mining can help to select the appropriate information (preprocessing), decrease the space of hypothesis (processing), to represent results in a more comprehensible way and to improve process (post processing)[3]. Ontology expresses the domain knowledge which includes semantic links between domain individuals described as relations of inter-concepts or roles [5]. Ontologies allow domain knowledge to be captured in an explicit and formal way such that it can be shared among human and computer systems. Figure 1 shows a concept hierarchy of food items based on the taxonomy presented in [25].

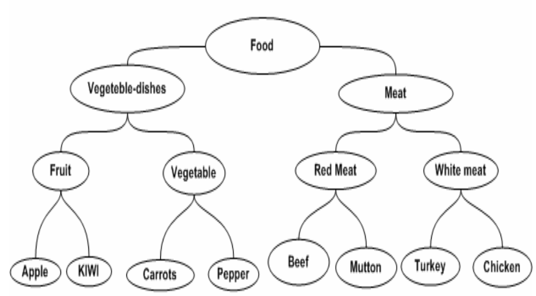

Figure 1: Crisp hierarchy example

2.2. Semantic distance in a Crisp ontology

Two main categories of algorithms for computing the semantic distance between terms organized in a hierarchical structure have been proposed in the literature [6]: distance-based approaches and information content-based approaches. The general idea behind the distance-based algorithms [20, 9, and 24] is to find the shortest path between two concepts in terms of number of edges. Information content-based approaches [7, 20] are inspired by the perception that pairs of words which share many common contexts are semantically related. We will be using distance-based approaches in this paper. 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 [20].

In the hierarchy of Figure 1, the edges distances between nodes of the graph are:
Dist(Apple, Kiwi) = 2 Dist(Carrots, Pepper) = 2
Dist(Apple, Meat) = 4 Dist(Fruit, Red Meat) = 4

2.3. Fuzzy Ontology

Knowledge about a system contains ambiguity and vagueness. It is convenient to represent the knowledge using fuzzy sets and fuzzy implications. The fuzzy ontology has been introduced to represent the fuzzy concepts and relationships where each concept is related to other concepts in the ontology, with a degree of membership μ ($0 \leq \mu \leq 1$) assigned to this 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.
In the example presented in Figure 2, Tomato may be regarded as being both Fruit and Vegetable, but to different degrees.
Our approach uses fuzzy membership degree in "is-a" relationships between concepts.

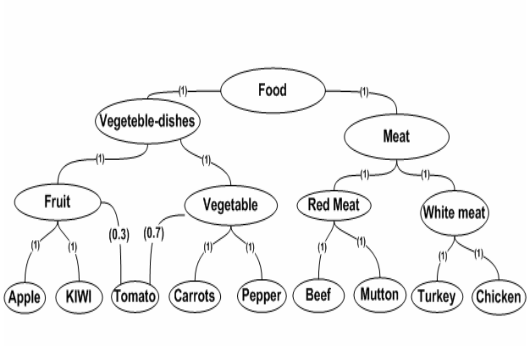

Figure 2: Fuzzy hierarchy example

2.4. Semantic distance in a fuzzy ontology

In the Crisp ontology as in Figure 1, the distance between two concept nodes, A and B, was defined as the shortest path that links A and B, i.e. the minimum number of edges that separate A and B or the sum of weights of the arcs along the shortest path between A and B [27].The semantic distance should be extended from the crisp case to the fuzzy case, considering that the weight of all edges is 1 in the crisp case. Rather than regarding fuzzy theory as a single theory, we should regard the process of "fuzzification" as a methodology to generalize any specific theory from a crisp (discrete) to a continuous (fuzzy) form [26].
To generalize the crisp weighting function, we propose an extension to the weighting function based on a boolean variable ($\mu \in \{0, 1\}$) to a weighting function based on a continuous variable ($\mu \in [0, 1]$).
The weighting function for the crisp hierarchy is $f(\mu)$: $\{0, 1\} \rightarrow \{0, 1\}$

$$\Rightarrow f(\mu) = \begin{cases} 1 & \text{if nodes } i, j \text{ connected} \\ 0 & \text{if nodes } i, j \text{ disconnected} \end{cases}$$

An extended versions to the preceding function is: $f(\mu)$: $[0,1] \rightarrow [0,2[$ where

$$f(\mu) = \begin{cases} 1 + \alpha(1-\mu) & \mu \neq 0 \\ 0 & \mu = 0 \end{cases}$$

α ($0 \leq \alpha \leq 1$) is weighting factor representing domain knowledge adjusted by an expert in the field.
μ ($0 \leq \mu \leq 1$) is the membership degree.

2.5. Degree of unexpectedness computation

For a given rule $R: X \rightarrow Y$ where $X = X_1 \wedge \ldots \wedge X_k$, $Y = Y_1 \wedge \ldots \wedge Y_m$ and D is the maximum depth of the hierarchy (D=3 for hierarchy in Figure 2), we define the degree of unexpectedness (DU) of a rule R as: DU(R)=Distance(X,Y)/4D.
To compute the distance between groups of concepts, we choose to use an extended version of the inter-concept distance measure of [20]:
$\text{Distance}(X_1 \wedge \ldots \wedge X_k, Y_1 \wedge \ldots \wedge Y_m) =$

$$\frac{1}{km} \sum_{i=1}^{k} \sum_{j=1}^{m} \text{Distance}(X_i, X_j)$$

This expression measures semantic distance between groups $X = X_1 \wedge \ldots \wedge X_k$ and $Y = Y_1 \wedge \ldots \wedge Y_m$ of concepts which contain k Xi and m atomic Yj concepts respectively.

2.6. Rules ranking

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

With such ranking, a user can check simply few patterns on the top of the list to confirm rule pertinence. The algorithm will be using the procedure to compute the weight of the edge based on the membership degree and using the proposed weighting function.

// this procedure (called by the algorithm), transform the membership degree to a weight for //each edge in the path and sum up the weights of the path going from Xi to Xj.

Procedure weight(ShortestPath(Xi,Xj))
Begin
 For X_k=X_i to X_j step 2
  Begin
   //Compute the weight of the edge.
   $w(X_k,X_{k+1}) = (1+ \alpha\ (1-\mu\ (X_k,X_{k+1}))$;
   //sum up the weight
   TotalWeight= TotalWeight +$w(X_k,X_{k+1})$
  End
 Return (TotalWeight)
End

Algorithm

ND: Number of nodes
R: Set of rules R= {Ri/ Ri=body→head}
  where i $\in$ [1,N]
N: number of rules
D: Maximum depth of the hierarchy
DU: Array of size N representing degree of unexpectedness
X_i, Y_j : Atomic Concepts; i $\in$ [1,k] ; j $\in$ [1,m]

Body = $X_1\wedge\ldots\wedge X_k$
Head = $Y_1\wedge\ldots\wedge Y_m$

//for all nodes in the graph calculate the semantic distance
For i=1 to ND
For j=1 to ND
Begin
 // ShortestPath(Xi,Xj) is the shortest path
//between the node Xi and the node Xj

 //Make call to weight(ShortestPath(Xi,Xj).
  Distance(Xi,Xj)=weight(ShortestPath(Xi,Xj))
End
//For each apply formula

$$\text{Dist}(X_1\wedge\ldots\wedge X_k, Y_1\wedge\ldots\wedge Y_m)=\frac{1}{km}\sum_{i=1}^{k}\sum_{j=1}^{m}\text{Dist}(X_i,X_j)$$

For i=1 to N
  DU[i] = (Dist($X_1\wedge\ldots\wedge X_k$; $Y_1\wedge\ldots\wedge Y_m$))/4D;

Sort Descending order DU.

3. Application

In this section we present results from applying our method to the hierarchy of Figure 2 for a set of association rules R = {Apple→ Kiwi; Apple→ Carrots; Pepper, Carrots →Turkey, Chicken; Kiwi →Tomato; Tomato → Pepper; Tomato, Pepper → Turkey, Chicken}

3.1. Nodes distance Computation

The number of graph nodes in (Figure 2) is ND=16 and the depth of the graph is D=3. For α=1, The weighting function $f(\mu) =1+(1-\mu)$.

The Weight between the nodes Tomato and fruit is:
 W(tomato,fruit)=1+(1-μ(tomato,fruit)) =1+(1-0.3) =1.7.

The weight W(tomato,Vegetable)= 1+(1-μ(tomato,Vegetable))=1+(1-0.7) =1.3

The semantic distance (the sum of the weight of the shortest path that separate 2 nodes) computation of Figure 2 graph nodes is presented in the following table (Table 1) where every cell represents the distance between the node on the line and the corresponding one on the column.

We have presented only the leaves of the hierarchy in Table 2 due to the fact that all the rules R are expressed using leaves concepts of the hierarchy.

	A	B	C	D	E	F	G	H	*I*
A	0	2	4	4	6	6	6	6	*2.7*
B	2	0	4	4	6	6	6	6	*2.7*
C	4	4	0	2	6	6	6	6	*2.3*
D	4	4	2	0	6	6	6	6	*2.3*
E	6	6	6	6	0	2	4	4	*6.3*
F	6	6	6	6	2	0	4	4	*6.3*
G	6	6	6	6	4	4	0	2	*6.3*
H	6	6	6	6	4	4	2	0	*6.3*
I	*2.7*	*2.7*	*2.3*	*2.3*	*6.3*	*6.3*	*6.3*	*6.3*	*0*

Table 1: Graph nodes distance

A=Apple ;B=Kiwi; C=Carrots; D=Pepper; E=Beef; F=Mutton; G=Turkey; H=Chicken; I=Tomato.

3.2. Rules degree of unexpectedness

The maximum depth of the hierarchy in (Figure 2) is D=3.

For a given rule X→Y where $X=X_1\wedge\ldots\wedge X_k$ and $Y=Y_1\wedge\ldots\wedge Y_m$

$$Distance(X,Y)=\frac{1}{km}\sum_{i=1}^{k}\sum_{j=1}^{m} Distance(X_i,X_j)$$

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

(a) Apple→ Kiwi

(b) Apple→ Carrots

(c) Pepper,Carrots →Turkey, Chicken

(d) Kiwi →Tomato

(e) Tomato → Pepper

(f) Tomato, Pepper → Turkey, Chicken

The detail computation distances of the rules (a), (b), (c), (d), (e) and (f) is:

(a) Dist(Apple, Kiwi)=2

(b) Dist (Apple, Carrots)=4

(c) Dist (Pepper $\wedge$ Carrots, Turkey $\wedge$ Chicken)=

1/4[Dist(Pepper,Turkey)+Dist(Pepper,Chicken)+Dist(Carrots,Turkey)+ Dist(Carrots, Chicken)] =

1/4[6+6+6+6]=6

(d) Dist(Kiwi, Tomato)=2.7

(e) Dist(Tomato, Pepper)=2.3

(f) Dist(Tomato $\wedge$ Pepper, Turkey $\wedge$ Chicken)=

1/4[Dist(Tomato,Turkey)+ Dist(Tomato,Chicken)+ Dist(Pepper,Turkey)+dist(Pepper, Chicken)] =

1/4[6.3+6.3+6+6]=36.6/4=6.15

The degree of unexpectedness for a given rule X→Y is calculated using our expression DU(X→Y)=Distance(X,Y)/4D and the resulting computation is presented in (Table 2).

Label	Rule	Dist	Degree Unexpectedness
(a)	Apple → Kiwi	2.00	2/12=0.16
(b)	Apple→Carrots	4.00	4/12=0.33
(c)	Pepper,Carrots→ Turkey,Chicken	6.00	6/12=0.50
(d)	Kiwi →Tomato	2.70	2.70/12=0.22
(e)	Tomato→Pepper	2.30	2.30/12=0.19
(f)	Tomato,Pepper→ Turkey, Chicken	6.15	6.15/12=0.51

Table 2: Rules degree of unexpectedness

The order of rules would be (f), (c), (b),(d),(e),(a) based on degree of unexpectedness descending order as shown in (Table 2). From decision system point of view the rules (f) and (c) belong to a higher level (Food) than the rules (b) and (d) that belongs to level (vegetable-dishes). The rule (e) and (a) belongs to a lower level (vegetable) and (Fruit) respectively. More we move up on in the hierarchy 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 domains (vegetables-dishes, Meat) which are farther than domains (vegetables, Fruits) of the rule (b) and (d). The rule (e) and (a) concerns only domain (vegetable) and (Fruit) respectively, and therefore they are less interesting. Note rule (d) is more surprising than rule (e) even though tomato is fruit and vegetable with different degree. The fact that tomato is closer to vegetable than fruit, the rule (d) is more interesting than the rule (e).

4. Related Works

Unexpectedness of patterns has been studied in [21, 22, 10, 11, 15, and 16] and defined in comparison with user beliefs. A rule is considered interesting if it affects the levels of conviction of the user. The unexpectedness is defined in probabilistic terms in [21, 22] while in [10] it is defined as a distance and it is based on a syntactic comparison between a rule and a conviction. Similarity and distance are 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 antecedents are distant or vice versa. In [17] 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 the discovered patterns. In [14] unexpectedness is defined from the point of view of a logical contradiction of a rule and conviction, the pattern that contradict a prior knowledge is unexpected. It is based on the contradiction of the consequence of the rule and the consequence of belief. Given a rule A→B and a belief X→Y, if B AND Y is False with A AND X is true for broad group of data, the rule is unexpected. In [12], 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. Most part of researches on the unexpectedness makes a syntactic or semantic comparison between a rule and a belief. Our definition of unexpectedness is based on the structure of background knowledge (hierarchy) underlying the terms (vocabulary) of the rule.

5. Conclusion and future work

In this paper we proposed a new approach for rule ranking according to their degree of unexpectedness, defined on the base of ontological distance. The ranking algorithm proposed uses a fuzzy ontology to calculate the distance between the antecedent and the consequent of rules on which is based the ranking. The more the conceptual distance is high, the more the rule represents a high degree of interest. We proposed a weighting function based on the membership degree to compute the weight of relations in the fuzzy ontology. This work constitutes a contribution to post analysis stage to help the user identify the most interesting patterns. In the future, we plan to incorporate a semantic distance threshold in the algorithm of calculation of frequent items, to exploit others relation of ontology other than “IS-A”.

6. References

[1] R. Agrawal, T. Imielinski, and A. Swami. Database mining: A performance perspective. IEEE Transactions on Knowledge and Data Engineering, 5(6):914–925, December 1993.

[2] Didier Dubois. Book review: ’fuzzy set theory and its applications’ (2nd edition) by H.J. zimmermann. Diffusion scientifique, 1991. BUSEFAL, Kluwer Academic Publ. Dordrecht, V.48, p.169-170.

[3] Zahra Farzanyar, Mohammadreza Kangavari, and Sattar Hashemi. A new algorithm for mining fuzzy association rules in the large databases based on ontology. In ICDM Workshops, pages 65–69. IEEE Computer Society, 2006.

[4] Usama M. Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. From data mining to knowledge discovery: An overview. In Advances in Knowledge Discovery and Data Mining, pages 1–34. 1996.

[5] T. R. Gruber. Towards principles for the design of ontologies used for knowledge sharing. In N. Guarino and

R. Poli, editors, Formal Ontology in Conceptual Analysis and Knowledge Representation, Deventer, The Netherlands, 1993. Kluwer Academic Publishers.

[6] Jay J. Jiang and David W. Conrath. Semantic similarity based on corpus statistics and lexical taxonomy. CoRR, cmp-lg/9709008, 1997. informal publication.

[7] Jay J. Jiang and David W. Conrath. Semantic similarity based on corpus statistics and lexical taxonomy. September 20 1997. Comment: 15 pages, Postscript only.

[8] Mika Klemettinen, Heikki Mannila, Pirjo Ronkainen, Hannu Toivonen, and A. Inkeri Verkamo. Finding interesting rules from large sets of discovered association rules. In Nabil R. Adam, Bharat K. Bhargava, and Yelena Yesha, editors, Third International Conference on Information and Knowledge Management (CIKM'94), pages 401–407. ACM Press, November 1994.

[9] Claudia Leacock and Martin Chodorow. Combining local context and WordNet similarity for word sense identification. In Christaine Fellbaum, editor, WordNet: An Electronic Lexical Database, pages 265–283. The MIT Press, Cambridge, Massachusetts, 1998.

[10] Bing Liu and Wynne Hsu. Post-analysis of learned rules. In Proceedings of the Thirteenth National Conference on Artificial Intelligence and the Eighth Innovative Applications of Artificial Intelligence Conference, pages 828–834, Menlo Park, August 4–8 1996. AAAI Press / MIT Press.

[11] Bing Liu, Wynne Hsu, and Shu Chen. Using general impressions to analyze discovered classification rules. In David Heckerman, Heikki Mannila, Daryl Pregibon, and Ramasamy Uthurusamy, editors, Proceedings of the Third International Conference on Knowledge Discovery and Data Mining (KDD-97), page 31. AAAI Press, 1997.

[12] Bing Liu,Wynne Hsu, Lai-Fun Mun, and Hing-Yan Lee. Finding interesting patterns using user expectations. IEEE Trans. Knowl. Data Eng, 11(6):817–832, 1999.

[13] John A. Major and John J. Mangano. Selecting among rules induced from a hurricane database. J. Intell. Inf. Syst, 4(1):39–52, 1995.

[14] B. Padmanabhan and A. Tuzhilin. On the discovery of unexpected rules in data mining applications. In On the Discovery of Unexpected Rules in Data Mining Applications. In Procs. of the Workshop on Information Technology and Systems (WITS '97), pages pp. 81–90, 1997.

[15] Balaji Padmanabhan and Alexander Tuzhilin. A belief-driven method for discovering unexpected patterns. In KDD, pages 94–100, 1998.

[16] Balaji Padmanabhan and Alexander Tuzhilin. Unexpectedness as a measure of interestingness in knowledge discovery, January 09 1999.

[17] Balaji Padmanabhan and Alexander Tuzhilin. On characterization and discovery of minimal unexpected patterns in rule discovery. IEEE Trans. Knowl. Data Eng, 18(2):202–216, 2006.

[18] Gregory Piatetsky-Shapiro. Discovery, analysis, and presentation of strong rules. In Knowledge Discovery in Databases, pages 231–233. AAAI/MIT Press, 1991.

[19] Gregory Piatetsky-Shapiro and Christopher J. Matheus. The interstigness of deviations. In KDD Workshop, pages 25–36, 1994.

[20] R. Rada, H. Mili, E. Bicknell, and M. Blettner. Development and application of a metric on semantic nets. IEEE Transactions on Systems, Man, and Cybernetics, 19(1):17–30, January-February 1989.

[21] Abraham Silberschatz and Alexander Tuzhilin. On subjective measures of interestingness in knowledge discovery. In KDD, pages 275–281, 1995.

[22] Abraham Silberschatz and Alexander Tuzhilin. What makes patterns interesting in knowledge discovery systems. IEEE Transactions on Knowledge and Data Engineering, 8(6):970–974, December 1996.

[23] Ramasamy Uthurusamy, Usama M. Fayyad, and W. Scott Spangler. Learning useful rules from inconclusive data. In Knowledge Discovery in Databases, pages 141–158. 1991.

[24] Zhibiao Wu and Martha Palmer. Verb semantics and lexical selection. In 32nd. Annual Meeting of the Association for Computational Linguistics, pages 133–138, New Mexico State University, Las Cruces, New Mexico, 1994.

[25] Chen, G., Wei, Q., and Kerre, E. E. (2000). Fuzzy Data Mining: Discovery of Fuzzy Generalized Association Rules. In Bordogna, G. and Pasi, G., editors, Recent Issues on Fuzzy Databases, pages 45–66. Physica-Verlag.

[26] L. A. Zadeh. Fuzzy sets. Information and Control, 8:338–353, 1965.

[27] Ray Richardson and Alan F. Smeaton. Using WordNet in a knowledge-based approach to information retrieval. Technical Report CA-0395, School of Computer Applications, Dublin City University, Dublin, Ireland, 1995.