

Ontology-Driven Method for Ranking Unexpected Rules

Mohamed Said Hamani¹ and Ramdane Maamri²

¹ Mohamed Boudiaf-M'sila University, Algeria saidhamani@hotmail.com

² 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. In this paper we propose a new approach for patterns ranking according to their unexpectedness using semantic distance calculated based on a prior background knowledge represented by domain ontology organized as DAG (Directed Acyclic Graph) hierarchy.

Keywords: data mining, ontology, unexpectedness, association rules, domain knowledge, subjective measures, semantic distance.

1 Introduction

Knowledge discovery in databases (data mining) has been defined in [6] 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 [23, 22, 30, 13]. 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 [23, 30, 13, 11]. Subjective measures of interestingness, such as unexpectedness [16, 31, 4], 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 [29, 30, 13, 14, 19, 20] where unexpectedness of discovered patterns is usually defined relative to a system of prior expectations. In this paper we define a degree of 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. 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 less 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 ontology organized as DAG hierarchy. The nodes of the hierarchy represent the rules vocabulary. For a rule like $(x \text{ AND } y \rightarrow z)$ x , y and z are nodes in the hierarchy. The semantic distance between the Antecedent $(x \text{ AND } y)$ 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 Semantic distance

Two main categories of algorithms for computing the semantic distance between terms organized in a hierarchical structure have been proposed in the literature [9]: distance-based approaches and information content-based approaches. The general idea behind the distance-based algorithms [24, 12, 32] is to find the shortest path between two terms in terms of number of edges. Information content-based approaches [10, 24] 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 elemental 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 or the sum of weights of the arcs along the shortest path between A and B [24].

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

2.2 Ontology

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)[5]. Ontology expresses the domain knowledge which includes semantic links between domain individuals described as relations of inter-concepts or roles [7].

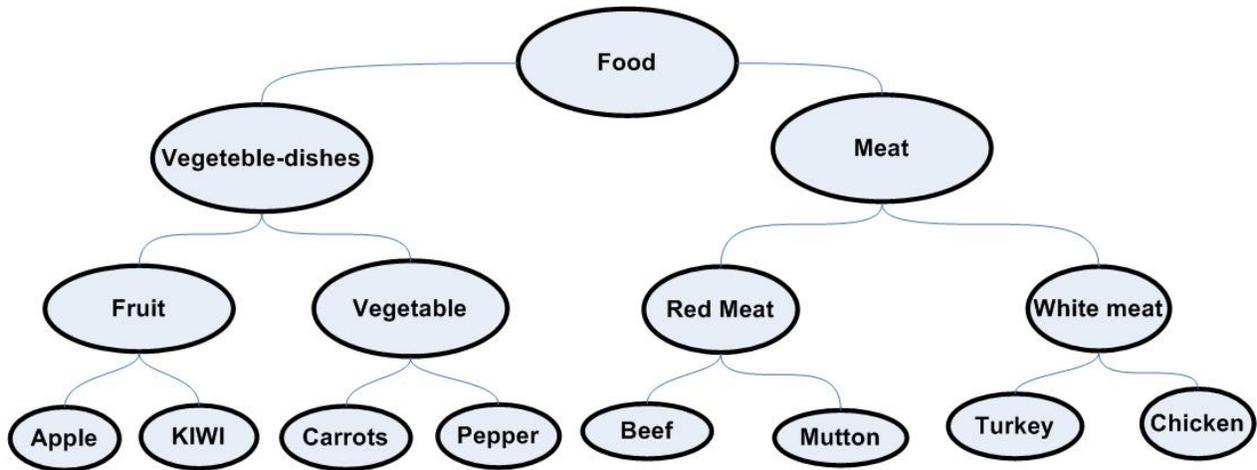


Fig. 1. hierarchy example

2.3 Degree of unexpectedness computation

For a given rule $R : X \rightarrow Y$ where $X = X_1 \wedge \dots \wedge X_k$, $Y = Y_1 \wedge \dots \wedge Y_m$ and D is the maximum depth of the hierarchy, we define the degree of unexpectedness (DU) of a rule R as : $DU(R) = \text{Distance}(X, Y) / 2D$.

To compute the distance between groups of concepts, we choose to use Hausdorff Distance

$\text{Distance}(X, Y) = \max(h(X, Y), h(Y, X))$ where

$$h(X, Y) = \max_{X_i \in X} \min_{Y_j \in Y} \|X_i - Y_j\|$$

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). It identifies the point $X_i \in X$ 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, as it reflects the distance of the point of X that is farthest from any point of Y and vice versa [8]. This expression measures semantic distance between groups $X_1 \wedge \dots \wedge X_k$ and $Y_1 \wedge \dots \wedge Y_m$ of concepts which contain k X_i and m atomic Y_j concepts respectively.

2.4 Rules ranking

In this section we introduce an algorithm to rank rules according to their degree of unexpectedness based on background knowledge. The rules we consider are on the form "body \rightarrow 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.

Algorithm

Input: Ontology, Set of rules

Output: Ordred set of rules

R: Set of rules $R = \{R_i / R_i = \text{body} \rightarrow \text{head}\}$ where $i \in [1, N]$

ND: Number of nodes

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 = $X_1 \wedge \dots \wedge X_k$

Head = $Y_1 \wedge \dots \wedge Y_m$

For $i=1$ to ND

begin

For $j=1$ to ND

Distance (X_i, X_j) =shortest path between X_i, X_j ;

End

For $i=1$ to N

begin

$DU[i] = (\text{Distance}(X_1 \wedge \dots \wedge X_k, Y_1 \wedge \dots \wedge Y_m))/2D$

End

Sort Descending degree of unexpectedness DU.

3 Example

In this section we present results from applying our method to the hierarchy of Figure 1 with a set of association rules $R = \{Apple \rightarrow Kiwi; Apple \rightarrow Carrots; Pepper, Carrots \rightarrow Turkey, Chicken\}$ resulting from a data mining process.

3.1 Nodes distance Computation

The number of graph nodes in Figure 1 is ND=16 and the depth of the graph is D=3.

The semantic distance (the minimum number of edges that separate 2 nodes) computation of Figure 1 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 1 due to the fact that all the rules R are expressed using leaves concepts of the hierarchy.

3.2 Degree of unexpectedness computation

The maximum depth of the hierarchy in Figure 1 is D=3.

For a given rule $X \rightarrow Y$ where $X = X_1 \wedge \dots \wedge X_k$ and $Y = Y_1 \wedge \dots \wedge Y_k$

Nodes	Apple	Kiwi	Carrots	Pepper	Beef	Mutton	Turkey	Chicken
Apple	0	2	4	4	6	6	6	6
Kiwi	2	0	4	4	6	6	6	6
Carrots	4	4	0	2	6	6	6	6
Pepper	4	4	2	0	6	6	6	6
Beef	6	6	6	6	0	2	4	4
Mutton	6	6	6	6	2	0	4	4
Turkey	6	6	6	6	4	4	0	2
Chicken	6	6	6	6	4	4	2	0

Table 1. Graph nodes distance

$$\text{Distance}(X,Y)= \text{Distance} (X_1 \wedge \dots \wedge X_k, Y_1 \wedge \dots \wedge Y_m) = \max(h(X, Y), h(Y, X))$$

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

- (a) Apple \rightarrow Kiwi
- (b) Apple \rightarrow Carrots
- (c) Pepper, Carrots \rightarrow Turkey, Chicken

The detail computation distance of the rules (a), (b), (c) is :

- (a) $\text{Dist}(\text{Apple}, \text{Kiwi})=2$
- (b) $\text{Dist}(\text{Apple}, \text{Carrots})=4$
- (c) $\text{Dist}(\text{Pepper} \wedge \text{Carrots}, \text{Turkey} \wedge \text{Chicken})=$
 $\max(h(\text{Pepper} \wedge \text{Carrots}, \text{Turkey} \wedge \text{Chicken}), h(\text{Turkey} \wedge \text{Chicken}, \text{Pepper} \wedge \text{Carrots}))$

$$h(\text{Pepper} \wedge \text{Carrots}, \text{Turkey} \wedge \text{Chicken})=6$$

$$h(\text{Turkey} \wedge \text{Chicken}, \text{Pepper} \wedge \text{Carrots})=6$$

$$(c) \text{Dist}(\text{Pepper} \wedge \text{Carrots}, \text{Turkey} \wedge \text{Chicken})= 6$$

The degree of unexpectedness for a given rule $X \rightarrow Y$ is calculated using our expression $\text{DU}(X \rightarrow Y) = \text{Distance}(X, Y) / 2D$ and the resulting computation is presented in (Table 2)

Label	Rule	Distance	Degree of unexpectedness
(a)	Apple \rightarrow Kiwi	2	2/6=0.33
(b)	Apple \rightarrow Carrots	4	4/6=0.66
(c)	Pepper, Carrots \rightarrow Turkey, Chicken	6	6/6=1.00

Table 2. Rules degree of unexpectedness

The order of rules would be (c), (b), (a) based on degree of unexpectedness descending order as shown in (Table 2). From decision system point of view the rule (c) belongs to a higher level (Food) than the rule (b) that belongs to level (vegetable-dishes). The rule (a) belongs to a lower level (Fruit). 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. Rule (c) is the crossing result of domains (vegetables-dishes, Meat) which are farther than domains (vegetables, Fruits) of the rule (b). The rule (a) concerns domain (Fruit) only and therefore it is the less interesting.

4 Experiments

The experiments were performed using a census income database with 48.842 records [3] with an implementation of our algorithm. To generate the association rules, we used the implementation of the Apriori algorithm [2] with a minimum support value equal 0.2 and a minimum confidence value equal 0.2. The number of the

generated rules set is 2225. In order to perform the experiments, we created the taxonomy of 81 weighted concepts based on the data set we are studying, as shown in (Table 3).

We conducted two tests, the first one with a weight value equals to one for all concepts with results presented in (Figure 2). The second test was conducted with different weights on the atomic concepts level (see Table 3 for weights), with results presented in (Figure 3). (Figure 2) and (Figure 3) are the extracted first two lines within each distance value for each test.

```
[Distance=6]      [Adm-clerical] <- [Female]
[Distance=6]      [Craft-repair] <- [HS-grad]
[Distance=6]      [Craft-repair] <- [EducationNum<=9]
[Distance=4]      [Exec-managerial] <- [>50K]
[Distance=4]      [Prof-specialty] <- [>50K]
```

Fig. 2. First Test Ranking Results

```
[Distance=12]     [Bachelors](7) <- [Never-married](1)
[Distance=12]     [Bachelors] <- [>50K, Married-civ-spouse]
[Distance=11]     [Craft-repair] <- [HS-grad]
[Distance=11]     [Bachelors] <- [>50K]
[Distance=9]      [Bachelors] <- [9<EducationNum<=13]
[Distance=9]      [Never-married] <- [Some-college]
[Distance=8]      [HS-grad] <- [>50]
[Distance=8]      [>50] <- [HS-grad]
[Distance=7]      [Craft-repair] <- [EducationNum<=9]
[Distance=7]      [Craft-repair] <- [30<age<=40, Male]
[Distance=6]      [Adm-clerical] <- [Female]
[Distance=6]      [9<EducationNum<=13] <- [Some-college]
[Distance=5]      [Private] <- [age>50]
[Distance=5]      [Private] <- [40<age<=50]
[Distance=4]      [Exec-managerial] <- [>50K]
[Distance=4]      [Prof-specialty] <- [>50K]
[Distance=3]      [Divorced] <- [40<age<=50]
[Distance=3]      [Divorced] <- [Female]
```

Fig. 3. Second Test Ranking Results

Looking to the results we notice:

1. Best results are those for highest weight (Figure 3 with Bachelors concept).
2. Best results from both tests are cross level concepts (higher subsumer like 'Personal', 'Education', 'Work' or 'census-income') and not those within the same concept level.
3. Low results from both tests (last 2 lines) are within the same concept level like 'Personal'.

Our approach is based on a hierarchy in (Table 3) which guides the resulting rules. The maximum hierarchy depth is 3 and it is the same as the minimum depth; this hierarchy is distributing the load equally between its different branches. The first test was conducted with weight equals to 1, for all concepts; In this case all

Level 0 Census-income	Level 1	Level 2	Atomic Concepts	weight
	work			1
		workclass		1
			Private	1
			Self-emp-not-inc	2
			Self-emp-inc	2
			Self-emp-inc	3
			Federal-gov	3
			Local-gov	5
			State-gov	5
			Without-pay	7
			Never-worked	7
		Occupation		1
			Tech-support	3
			Craft-repair	3
			Other-service	1
			Sales	1
			Exec-managerial	2
			Prof-specialty	2
			Handlers-cleaners	2
			Machine-op-inspct	2
			Adm-clerical	2
			Farming-fishing	2
			Transport-moving	2
			Protective-serv	2
			Armed-Forces	1
		SalaryClass		1
			$\leq 50K$	0
			$> 50K$	0
	Education			1
		educationLevel		1
			Tech-support	4
			Bachelors	7
			Some-college	4
			11th	4
			HS-grad	4
			Prof-school	4
			Assoc-acdm	4
			.	
			.	
			Preschool	6
		EducationNum		1
			≤ 9	0
			$9 < num \leq 13$	0
			$13 < num \leq 15$	0
			$num > 15$	0
	Personal			1
		race		1
			White	0
			Black	0
			Asian-Pac-Islander	0
			Amer-Indian-Eskimo	0
			Other	0
		sex		1
			Female	0
			Male	0
		native-country		1
			Europe	0
			Asia	0
			SouthAmerica	0
			United-States	0
			Canada	0
		age		1
			$age \leq 20$	0
			$20 < age \leq 30$	0
			$30 < age \leq 40$	0
			$40 < age \leq 50$	0
			> 50	0
		marital-status		1
			Married-civ-spouse	1
			Divorced	1
			Never-married	1
			Separated	1
			Widowed	1
			Married-spouse-absent	3
			Married-AF-spouse	3

Table 3. Experiment Taxonomy

concepts have the same degree of interest to the user. The ranking rules algorithm picks those with higher subsumer concept. The common subsumer for the rules ((1), (2) and (3) of (Figure 2) is the top concept 'census-income', however The common subsumer for the rules (4) and (5) is the concept 'Work'. Rule (1) concerns 'sex' and 'occupation', however rules (2) and (3) are about education and occupation. The last 2 rules (4) and (5) express the relation between 'occupation' and 'salary-class'. We believe a rule like (1), (2) or (3) is more interesting, because it is giving us information between 'Education' and 'Personal' information and it involves a higher decision maker (strategic) than the one concerning 'occupation' and 'salary' that can concerns payroll for instance.

The second test was conducted with a weight of 'bachelors' concept equals to 7 (among other concepts settings see Table 3).The user in this case is putting more emphasis on this concept by setting its weight to a high value or because it is really that important in the domain of study. The ranking rules algorithm picks those with higher weight. The common subsumer for the rules (1) and (2) of (Figure 3) is the concept 'census-income', but in this case with a 'Bachelors' concept as member of the rule. In this case the user is focusing his study on people with 'bachelors' education and their relation to 'Personal' information or 'Work'. The common subsumer for the last 2 rules of (Figure 3) is the concept 'Personal'. These rules express the relation between 'sex', 'age' and 'matrial-status' concepts. Even though interestingness is subjective (What's interesting of one may not be of the same degree of interest to the other), we believe more we move up on in the hierarchy, more the decision is important and the vision of the decision maker is broader, strategic and important; therefore the discovered rule is more interesting. Our approach follows this vision.

5 Related Works

Unexpectedness of patterns has been studied in [29, 30, 13, 14, 19, 20] 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 [29, 30] while in [13] 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 [21] 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 [18] 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 $A \rightarrow B$ and a belief $X \rightarrow Y$, if $B \text{ AND } Y$ is False with $A \text{ AND } X$ is true for broad group of data, the rule is unexpected. In [15], 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. [26, 27, 28] has taken a different approach to the discovery of interesting patterns by eliminating noninteresting association rules. Rather than getting the users define their entire knowledge of a domain, they are asked to identify several non-interesting rules, generated by the Apriori algorithm. [25] use genetic algorithm to dynamically maintain and search populations of rule sets for the most interesting rules rather than act as post-processor. The rules identified by the genetic algorithm compared favorably with the rules selected by the domain expert [17]. Most 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. We are taking a different approach from all the preceding work. The preceding work is a filtering process based on a belief expressed as rules that the user has to enter. We are proposing a ranking process and the knowledge are not expressed as rules, but as hierarchy of concepts ontology. Ontologies enable knowledge sharing. Sharing vastly increases the potential for knowledge reuse and therefore allows our approach to get free knowledge just from using domain ontologies already available like "ONTODerm" for dermatology, "BIO-ONT" for biomedicine, "ASFA, OneFish, FIGIS, AGROVOC" for Food, etc.

6 Conclusion and future work

In this paper we proposed a new approach to estimate the degree of unexpectedness of a rule with respect to ontology and ranking patterns according to their unexpectedness, defined on the base of ontological distance. The ranking algorithm proposed uses an 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. 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". We are also validating our approach on fuzzy ontology to take into account vague and imprecise information.

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] Chistian Borgelt. <http://www.borgelt.net/software.html>.
- [3] census income. <ftp://ftp.ics.uci.edu/pub/machine-learning-databases/census-income/>.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] Daniel P. Huttenlocher, Gregory A. Kl, and William J. Rucklidge. Comparing images using the hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15:850–863, 1993.
- [9] Jay J. Jiang and David W. Conrath. Semantic similarity based on corpus statistics and lexical taxonomy. *CoRR*, cmp-lg/9709008, 1997. informal publication.
- [10] Jay J. Jiang and David W. Conrath. Semantic similarity based on corpus statistics and lexical taxonomy. September 20 1997. Comment: 15 pages, Postscript only.
- [11] 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.
- [12] 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.
- [13] 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, 1996. AAAI Press / MIT Press.
- [14] 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.
- [15] 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.
- [16] John A. Major and John J. Mangano. Selecting among rules induced from a hurricane database. *J. Intell. Inf. Syst.*, 4(1):39–52, 1995.
- [17] Kenneth McGarry. A survey of interestingness measures for knowledge discovery. *Knowledge Eng. Review*, 20(1):39–61, 2005.
- [18] 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.

- [19] Balaji Padmanabhan and Alexander Tuzhilin. A belief-driven method for discovering unexpected patterns. In *KDD*, pages 94–100, 1998.
- [20] Balaji Padmanabhan and Alexander Tuzhilin. Unexpectedness as a measure of interestingness in knowledge discovery, January 09 1999.
- [21] 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.
- [22] Gregory Piatetsky-Shapiro. Discovery, analysis, and presentation of strong rules. In *Knowledge Discovery in Databases*, pages 231–233. AAAI/MIT Press, 1991.
- [23] Gregory Piatetsky-Shapiro and Christopher J. Matheus. The interstigness of deviations. In *KDD Workshop*, pages 25–36, 1994.
- [24] 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.
- [25] Wesley Romão, Alex Alves Freitas, and Itana Maria de Souza Gimenes. Discovering interesting knowledge from a science and technology database with a genetic algorithm. *Appl. Soft Comput.*, 4(2):121–137, 2004.
- [26] Sigal Sahar. Interestingness via what is not interesting. In *KDD*, pages 332–336, 1999.
- [27] Sigal Sahar. Interestingness preprocessing. In *ICDM*, pages 489–496, 2001.
- [28] Sigal Sahar. On incorporating subjective interestingness into the mining process. In *ICDM*, pages 681–684, 2002.
- [29] Abraham Silberschatz and Alexander Tuzhilin. On subjective measures of interestingness in knowledge discovery. In *KDD*, pages 275–281, 1995.
- [30] Abraham Silberschatz and Alexander Tuzhilin. What makes patterns interesting in knowledge discovery systems. *IEEE Transactions on Knowledge and Data Engineering*, 8(6):970–974, 1996.
- [31] Ramasamy Uthurusamy, Usama M. Fayyad, and W. Scott Spangler. Learning useful rules from inconclusive data. In *Knowledge Discovery in Databases*, pages 141–158. 1991.
- [32] 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.