A Conformity Measure using Background Knowledge for Association Rules: Application to Text Mining

Hacène Cherfi, Amedeo Napoli, Yannick Toussaint

A Conformity Measure using Background Knowledge for Association Rules: Application to Text Mining

Hacène Cherfi
INRIA Sophia Antipolis, 06902 Sophia Antipolis, France
hacene.cherfi@sophia.inria.fr
Amedeo NAPOLI
LORIA - INRIA, 54506 Vandoeuvre-lès-Nancy, France
amedeo.napoli@loria.fr
Yannick TOUSSAINT
LORIA - INRIA, 54506 Vandoeuvre-lès-Nancy, France
yannick.toussaint@loria.fr

ABSTRACT

A text mining process using association rules generates a very large number of rules. According to experts of the domain, most of these rules basically convey a common knowledge, i.e. rules which associate terms that experts may likely relate to each other. In order to focus on the result interpretation and discover new knowledge units, it is necessary to define criteria for classifying the extracted rules. Most of the rule classification methods are based on numerical quality measures. In this chapter, we introduce two classification methods: The first one is based on a classical numerical approach, i.e. using quality measures, and the other one is based on domain knowledge. We propose the second original approach in order to classify association rules according to qualitative criteria using domain model as background knowledge. Hence, we extend the classical numerical approach in an effort to combine data mining and semantic techniques for post mining and selection of association rules. We mined a corpus of texts in molecular biology and present the results of both approaches, compare them, and give a discussion on the benefits of taking into account a knowledge domain model of the data.

KEYWORDS

Knowledge extraction, Text Mining, Association Rules, Semantic Measure, Statistical Measure, Domain knowledge model, Natural language processing.

1. INTRODUCTION

From the data mining point of view, texts are complex data giving raise to interesting challenges. First, texts may be considered as weakly structured, compared with databases that rely on a predefined schema. Moreover, texts are written in natural language, carrying out implicit knowledge, and ambiguities. Hence, the representation of the content of a text is often only partial and possibly noisy. One solution for handling a text or a collection of texts in a satisfying way is to take advantage of a knowledge model of the domain of the texts, for guiding the extraction of knowledge units from the texts.

In this chapter, we introduce a knowledge-based text mining process (KBTM) relying on the knowledge discovery process (KDD) defined in [Fayyad *et al.*, 1996]. The KBTM process relies on an interactive loop, where the analyst – an expert of the text domain – controls and guides the mining process. The objective of the mining process is to enrich the knowledge model of the text domain, and, in turn, to improve the capability of the knowledge-based text mining process itself.

Following a natural language processing of the texts described in [Cherfi *et al.*, 2006], the text mining process (also denoted by TM in the following) is applied to a binary table Texts × Keyterms, and produces a set of association rules (AR in the following). The set Keyterms includes a set of keyterms giving a kind of summary of the content of each text. The extraction of association rules is carried out thanks to a *frequent itemse*t algorithm (namely the Close algorithm [Pasquier *et al.*, 1999]). Association rules show some advantages, among which the facts that AR are easily understandable and that they highlight regularities existing within the set of texts.

Two text mining approaches based on association rules are studied hereafter. The first approach is based on the use of statistical quality measures for classifying the extracted rules [Cherfi et al., 2006]. A set of five quality measures is introduced, each of them expressing some particular aspects of the texts: e.g. rare keyterms, functional dependencies, or probabilistic correlations between keyterms. One limitation of this approach is due to the numerical characteristics of the classification process, which takes into account the distribution of the keyterms, and ignores the semantics carried by the keyterms. By contrast, a second approach is based on a domain knowledge model of the texts which is used to classify the extracted association rules. The knowledge model is a pair $(K, \sqsubseteq)$ where K is a finite set of keyterms and $\sqsubseteq$ is a specialisation relation (*i.e.*, a partial ordering). Hence, the quality of a rule depends on the conformity of the rule with respect to the knowledge model: a rule is interesting if it includes semantic relations that are not already known in the knowledge model. Thus, the knowledge model is used to guide the interpretation and the classification of the extracted association rules. This KBTM approach is original and relies on a qualitative approach rather than on a more classical approach based on statistical quality measures. Two experiments show that the KBTM approach gives substantial and good quality results, opening new perspectives in the difficult field of text mining. The objective of these experiments is to show how far our proposed Conformity measure is consistent with the text mining task in a specific domain (here molecular biology).

This chapter is organised as follows. Firstly, we introduce the context of association rule extraction for text mining, and we present and discuss an example, based on statistical quality measures. Then, we introduce the principles of the KBTM process. We analyse thanks to an example –the same as in the first part of the chapter – the KBTM process for the so-called simple and complex extracted AR. The following section sets up an experiment and a qualitative analysis based on real-world collection of texts with the help of an analyst. The AR are classified according to the conformity measure, in contrast with five statistical measure classifications. We continue the chapter with a discussion on the benefits of the KBTM approach, and we mention some related work. The chapter ends with a conclusion and draws future work.

2. EXTRACTION OF ASSOCIATION RULES FOR TEXT MINING

2.1. Text processing for data mining preparation

In our experiments, we dealt with a collection of texts (hereafter called corpus) in molecular biology. Basically, we start with a set of bibliographical records characterised by contextual metadata, *e.g.*, title, author(s), date, status (whether published or not), keywords, etc. Hereafter, we explain how we get the keyterms associated with each text.

Extracting Textual Fields in the Sources: A first processing of this collection of records consists in extracting two textual fields, the title and the abstract.

Part-of-speech (POS) Tagging: It is a natural language processing (NLP) technique which associates with each word of the texts a linguistic tag corresponding to its grammatical category (noun, adjective, verb, etc.). A POS-tagger needs a learning phase with a manually tagged vocabulary. A POS-tagger basically uses a statistical model to learn how to predict the category of a word with respect to the preceding word categorisation. Several taggers exist for English and show high performance of correctness [Paroubek, 2007]. For example, sentence (1) extracted from one of our texts gives the tagged sentence (2):

1. Two resistant strains were isolated after four rounds of selection.
2. Two/CD resistant/JJ strains/NNS:pl were/VBD isolated/VBN after/IN four/CD rounds/NNS:pl of/IN selection/NN.

Terminological Indexing: In our experiments, the texts have been processed and represented by a set of keyterms. A keyterm is a noun phrase (*i.e.*, one to many words) of our vocabulary which can be associated with a domain concept of our knowledge model, thus, it ensures the transition from the linguistic to the knowledge level.

Keyterm Identification and Variants: We have used the FASTR [Jacquemin, 1994] terminological extraction system for identifying the keyterms of our vocabulary in the text. It allows us to recognise a keyterm in several variant forms. For example, the expression “transfer of capsular biosynthesis genes” is considered as a variant form of the keyterm “gene transfer” which belongs to the vocabulary. However, all the variants are not acceptable; NLP meta-rules are used to keep the variants preserving the initial sense of the keyterm. The keyterm variants are identified using the meta-rules. A meta-rule is a transformation rule operating on the grammatical description of a keyterm and the linguistically authorised variation of this description. For example, the expression “transfer of genes” is recognised as a variation of the keyterm “gene transfer” (which belongs to the vocabulary) by a *permutation* meta-rule of “gene” and “transfer”. The expression “transfer of capsular biosynthesis genes” is recognised as well by applying an *insertion* meta-rule (of “capsular biosynthesis”). In this way, the NLP keyterm identification contributes to reduce the word dispersion in the description of a text by unifying variants to a single keyterm.

2.2. Association Rules and Statistical Quality measures

Let $T = \{t_1 , t_2, ..., t_m\}$ be a set of m texts and $K = \{k_1 , k_2 , ... , k_n\}$ a set of n keyterms associated with these texts. An association rule is a weighted implication such as $A \rightarrow B$ where $A = \{k_1 , k_2, ..., k_p\}$ (the *body*) and $B = \{k_{p+1} , k_{p+2}, ..., k_q\}$ (the *head*). The rule $A \rightarrow B$ means that if a text contains $\{k_1 , k_2, ..., k_p\}$ then it tends to contain also $\{k_{p+1} , k_{p+2}, ..., k_q\}$ with a probability given by the confidence of the rule. Several algorithms aim at extracting association rules: Apriori [Agrawal *et al.*, 1996] or Close [Pasquier et *al.*, 1999] that will be used hereafter. The support and the confidence are two quality measures related to association rules that are used to reduce the number of the extracted units, hence reducing the complexity of the extraction process. The support of a rule $A \rightarrow B$ measures the number of texts containing both keyterms of A and B. The union of the keyterm sets A and B is denoted by $A \sqcap B$. The support may be normalised by the total number of texts. The confidence of a rule is defined by the ratio between the number of texts containing the keyterms in $A \sqcap B$, and the number of texts containing the keyterms in A. The confidence is seen as the conditional probability $P(B/A)$. The confidence of a rule measures the proportion of examples and counterexamples of the rule. A counterexample states that there exist texts having all the keyterms of A, but not necessarily all the keyterms of B. When the confidence of a rule is 1, the rule is *exact*, otherwise it is *approximate*. Two thresholds are defined, σ_s for the minimum support, and σ_c for the minimum confidence. A rule is valid whenever its support is greater than σ_s and its confidence is greater than σ_c.

Considering a rule such as $A \rightarrow B$, if A and B are frequent keyterm sets (*i.e.*, their support is above the σ_s threshold), then they are shared by a large proportion of texts, and the probabilities $P(A)$, $P(B)$, and $P(A \sqcap B)$ are high (here probability stands for the number of texts containing a given keyterm set out of the total number of the texts). The importance of such frequent keyterm sets is rather small, from the KDD point of view. By contrast, when A and B are rare, i.e. they have a low probability, then these keyterm sets are shared by a low number of texts, i.e. the keyterms in A and B may be related in the context of the mined text set. However, the support and the confidence are not always sufficient for classifying extracted association rules in a meaningful way. This reason leads to introduce a number of other statistical quality measures attached to the rules enlightening some particular aspects on the rules [Lavrac *et al.*,1999]. Five of these quality measures are presented hereafter, and have been used in our two experiments.

1. The *interest* measures the degree of independence of the keyterm sets A and B, and is defined by $\text{interest}(A \rightarrow B) = P(A \sqcap B)/P(A) \times P(B)$. The interest is symmetrical ($\text{interest}(A \rightarrow B = \text{interest}(A \rightarrow B)$) and has its range in the interval $[0, +\infty[$. It is equal to 1 whenever the "events" A and B are statistically independent. The more A and B are incompatible, the more $P(A \sqcap B)$, and hence the interest, tend to 0;
2. The *conviction* allows us to select among the rules $A \rightarrow B$ and $A \rightarrow B$ the one having the less counterexamples. The conviction is defined by $\text{conviction}(A \rightarrow B) = P(A) \times P(\neg B)/P(A \sqcap \neg B)$. The conviction is not symmetrical, and has its range in $[0, +\infty[$. It denotes a dependency between A and B whenever it is greater than 1, independence whenever it is equal to 1, and no dependency at all whenever it is lower than 1. The conviction is not computable for exact rules because $P(A \sqcap \neg B)$ is equal to 0 (there is no counterexample for exact rules);
3. The *dependency* measures the distance between the confidence of the rule and the independence case: $\text{dependency}(A \rightarrow B) = |P(B/A) - P(B)|$. This measure has its range in $[0, 1[$, where a

dependency close to 0 (respectively to 1) means that A and B are independent (respectively dependent);

4. The *novelty* is defined by novelty(A → B) = P(A ⊓ B) − P(A) × P(B), and has its range within]−1, 1[, with a negative value whenever P(A ⊓ B) < P(A) × P(B). The novelty tends to −1 for rules with a low support, i.e. P(A ⊓ B) ≃ 0. The novelty is symmetrical although the rule A → B may have more counterexamples than the rule B → A. It leads to the definition of the following measure;
5. The *satisfaction* measure is defined by satisfaction(A → B) = P(¬B)− P(¬B|A)/P(¬B) . The satisfaction has its range in [−∞, 1], and is equal to 0 whenever A and B are independent. The satisfaction cannot be used for classifying exact rules because, in this case, its value is equal to 1.

2.3. Using Quality Measures on a Small Example

An example borrowed from [Pasquier *et al.*, 1999] will be used to illustrate the behaviour of the statistical quality measures introduced above. Let us consider six texts {t_1, t_2, t_3, t_4, t_5, t_6} described by a set of five keyterms, namely {a, b, c, d, e}. So the text t1 is described by the keyterm set {b, c, e} (see Table 1), and hereafter more simply denoted by the bce. The extraction of the association rules has been performed with the Close algorithm [Pasquier *et al.*, 1999]. Twenty association rules, numbered r_1 , . . . , r_{20} , have their support greater than the threshold σ_s = 1/6 (where 6 is the total number of texts), and their confidence is greater than σ_c = 0.1 (or 10%). The set of extracted association rules is given in Table 2. The rules have been extracted from closed frequent keyterm sets. The Close algorithm is based on levelwise search of *closed frequent* keyterm sets in the binary table Texts × Keyterms, starting from the smallest closed keyterm sets {ac, be} to the largest closed keyterm set abce. A closed frequent keyterm set corresponds to a maximal set of keyterms shared by a given subset of texts, with a support greater than the σ_s threshold. Once the closed frequent keyterm sets have been extracted, the association rules of the form $P_2 \rightarrow P_1 \setminus P_2$ may be derived, where for example b → ce stands for "b → bce \ b". The extracted association rules A → B have a minimal *body*, i.e. A corresponds to a generator, and a maximal head, i.e. B corresponds to a closed set for the Galois connection associated with the relation Texts × Keyterms (see for example [Bastide *et al.*, 2000]). For example, the association rules b → e and b → c ∧ e are extracted, because the corresponding keyterm sets be and bce are closed sets in the Galois connection.

Table 1.The textual database

Texts	Keyterms
t_1	acd
t_2	bce
t_3	abce
t_4	be
t_5	abce
t_6	bce

Table 2.The set of 20 valid AR

id	Rule	id	Rule
r_1	$b \rightarrow e$	r_{11}	$a \rightarrow c$
r_2	$b \rightarrow c \wedge e$	r_{12}	$b \wedge c \rightarrow a \wedge e$
r_3	$a \wedge b \rightarrow c \wedge e$	r_{13}	$d \rightarrow a \wedge c$
r_4	$a \rightarrow b \wedge c \wedge e$	r_{14}	$c \rightarrow b \wedge e$
r_5	$b \wedge c \rightarrow e$	r_{15}	$c \rightarrow a \wedge d$
r_6	$b \rightarrow a \wedge c \wedge e$	r_{16}	$c \rightarrow a \wedge b \wedge e$
r_7	$e \rightarrow b \wedge c$	r_{17}	$c \wedge e \rightarrow b$
r_8	$a \wedge e \rightarrow b \wedge c$	r_{18}	$c \wedge e \rightarrow a \wedge b$
r_9	$a \rightarrow c \wedge d$	r_{19}	$e \rightarrow b$
r_{10}	$e \rightarrow a \wedge b \wedge c$	r_{20}	$c \rightarrow a$

The classification of the rules according to the different quality measures is given in Table 3. In each column of the table, the rules are classified according to the value of the measure in a decreasing order. Such a rule classification may be presented to an analyst, either for the whole set of measures or only one particular measure. An algorithm for classifying extracted association rules according to these quality measures (and their roles) is proposed in [Cherfi *et al.*, 2006].

Table3.Statistical measures for the 20 valid AR in a decreasing order

id	support	id	confidence	id	interest	id	conviction	id	dependence	id	novelty	id	satisfaction
r_1	5	r_1	1.000	r_9	2.000	r_7	1.667	r_{13}	0.500	r_1	0.139	r_1	1.000
r_2	5	r_3	1.000	r_{13}	2.000	r_2	1.667	r_3	0.333	r_{19}	0.139	r_3	1.000
r_6	5	r_5	1.000	r_3	1.500	r_{12}	1.333	r_8	0.333	r_2	0.111	r_5	1.000
r_7	5	r_8	1.000	r_8	1.500	r_{18}	1.333	r_1	0.167	r_3	0.111	r_8	1.000
r_{10}	5	r_{11}	1.000	r_{12}	1.500	r_9	1.250	r_5	0.167	r_5	0.111	r_{11}	1.000
r_{14}	5	r_{13}	1.000	r_{18}	1.500	r_{20}	1.250	r_9	0.167	r_7	0.111	r_{13}	1.000
r_{15}	5	r_{17}	1.000	r_1	1.200	r_6	1.111	r_{11}	0.167	r_8	0.111	r_{17}	1.000
r_{16}	5	r_{19}	1.000	r_2	1.200	r_{10}	1.111	r_{12}	0.167	r_9	0.111	r_{19}	1.000
r_{19}	5	r_2	0.800	r_5	1.200	r_{16}	1.111	r_{17}	0.167	r_{11}	0.111	r_2	0.400
r_{20}	5	r_7	0.800	r_6	1.200	r_{15}	1.042	r_{18}	0.167	r_{12}	0.111	r_7	0.400
r_5	4	r_{14}	0.800	r_7	1.200	r_4	1.000	r_{19}	0.167	r_{13}	0.111	r_{12}	0.250
r_{12}	4	r_4	0.667	r_{10}	1.200	r_{14}	0.833	r_2	0.133	r_{17}	0.111	r_{18}	0.250
r_{17}	4	r_{20}	0.600	r_{11}	1.200	r_1	0.000	r_7	0.133	r_{18}	0.111	r_9	0.200
r_{18}	4	r_{12}	0.500	r_{15}	1.200	r_3	0.000	r_{20}	0.100	r_{20}	0.111	r_{20}	0.200
r_4	3	r_{18}	0.500	r_{16}	1.200	r_5	0.000	r_6	0.067	r_6	0.056	r_6	0.100
r_9	3	r_6	0.400	r_{17}	1.200	r_8	0.000	r_{10}	0.067	r_{10}	0.056	r_{10}	0.100
r_{11}	3	r_{10}	0.400	r_{19}	1.200	r_{11}	0.000	r_{16}	0.067	r_{16}	0.056	r_{16}	0.100
r_3	2	r_{16}	0.400	r_{20}	1.200	r_{13}	0.000	r_{14}	0.033	r_{15}	0.028	r_{15}	0.040
r_8	2	r_9	0.333	r_4	1.000	r_{17}	0.000	r_{15}	0.033	r_4	0.000	r_4	0.000
r_{13}	1	r_{15}	0.200	r_{14}	0.960	r_{19}	0.000	r_4	0.000	r_{14}	-0.028	r_{14}	-0.200

3. CONFORMITY OF AN ASSOCIATION RULE WITH RESPECT TO A KNOWLEDGE MODEL

3.1. Conformity for a Simple Rule

Definition 1 (Knowledge Model)

A knowledge model, denoted by $(K, \sqsubseteq)$, is a finite, directed graph with K standing for the set of vertices (the keyterms), and the relation $\sqsubseteq$ defining the edges of the graph and the partial ordering over the keyterms in K. For each $x, y \in K$, $x \sqsubseteq y$ means that each instance of the keyterm concept x is also an instance of the keyterm concept y.

The principle of classifying AR according to their conformity with a knowledge model is stated as follows: we assign a high value of conformity to any association rule $A \rightarrow B$ that is "represented" in $(K, \sqsubseteq)$ with a relation $A \sqsubseteq B$ existing between the keyterms $a_i \in A$ and $b_j \in B$, $i, j \geq 1$. We suppose in the following of this section that the rules are simple in the sense that their *body* and *head* are restricted to a single keyterm, for example $b \rightarrow e$. The so-called complex rules where the *body* and/or the *head* are composed of more than one keyterm are considered in section 3.4.

Definition 2 (Conformity for a Simple AR with the Knowledge Model)

Let k_1, k_2 be in K, and let $k_1 \rightarrow k_2$ be a valid AR. The conformity measure of $k_1 \rightarrow k_2$ with $(K, \sqsubseteq)$ is defined by the probability of finding out a path from k_1 to k_2 – called hereafter the probability transition from k_1 to k_2 – in the directed graph of $(K, \sqsubseteq)$. This path can be composed of one to several edges.

If we consider that updating the knowledge model consists in introducing new keyterms and new relations between keyterms in K, then an association rule $x \rightarrow y$ is conform to $(K, \sqsubseteq)$ (i.e., it has a high value of conformity) if the relation $x \sqsubseteq y$ exists in $(K, \sqsubseteq)$. Otherwise, the rule is not conform to the knowledge model (i.e., its conformity value is low). Indeed, we have to notice that a rule $x \rightarrow y$ extracted within the text mining process is not added to $(K, \sqsubseteq)$ without the control of the analyst in charge of updating a knowledge model of his domain. Any knowledge unit update is supervised by the analyst. The computation of the conformity is based on the principles of the spreading activation theory [Collins & Loftus,1975] stating that the propagation of an information marker through the graph of the knowledge model from a given vertex, say k_1 , to another vertex, say k_2 , relies on the strength associated to the marker. The value of the strength depends on: (i) the length of the path, and (ii) on the number of reachable keyterms starting from k_1 in $(K, \sqsubseteq)$. The strength of the marker monotonically decreases with respect to these two factors.

Definition 3 (Calculation of the Conformity Measure for Simple Rules)

The conformity of a simple rule $k_1 \rightarrow k_2$ is defined as the transition probability from the keyterm k_1 to the keyterm k_2 , and is dependent on the minimal path length between k_1 and k_2 , and the centrality of k_1 in $(K, \sqsubseteq)$ which depends on how many keyterms are related to k_1 in $(K \setminus k_1)$.

3.2. Transition Probability

Given the domain knowledge model $(K, \sqsubseteq)$, a probability transition table is set and used as a basis of the conformity calculation. The probability transition of k_i and k_j depends on the minimal distance $d(k_i, k_j)$ between a keyterm k_i and a keyterm k_j in $(K, \sqsubseteq)$. We distinguish two particular cases:

1. For each k_i, $d(k_i,k_i) = 1$ in order to take into account the reflexivity of the relation $\sqsubseteq$, and to avoid abnormally high probabilities in a case where there is no outgoing edge from k_i (as illustrated by the vertex c in Figure 1);
2. If it does not exist a path from a keyterm k_i to a keyterm k_j , then we set a "minimal" (non zero) transition probability by using $d(k_i,k_j) = 2N+1$, where N is the cardinal of the set of keyterms in K.

The transition probability from k_i to k_j , denoted by $Cty(k_i,k_j)$, defines the Conformity measure of the rule $k_i \rightarrow k_j$, and relies on the product of two elements: (i) the distance from k_i to k_j , and (ii) a normalisation factor, denoted by $\delta(k_i)$. Moreover, two additional principles are used:

1. The higher the distance between two keyterms k_i and k_j is, the lower the conformity for $k_i \rightarrow k_j$ is;
2. The normalisation factor of a keyterm k_i depends on all the keyterms in K, either they are reachable from k_i or not. Putting things altogether, the formula for calculating the conformity for a simple rule is stated as follows: $\boxed{Cty(k_i,k_j) = [d(k_i,k_j) \times \delta(k_i)]^{-1}}$ where the normalisation factor of k_i is: $\delta(k_i) = \Sigma_{x \in K} 1/d(k_i,x)$.

Hence, $\delta(k_i)$ depends on the number of outgoing edges from k_i in K: the higher the number of outgoing edges from k_i is, the lower $\delta(k_i)$ is. In accordance, when there is no outgoing edge from a keyterm k_i; this keyterm k_i becomes "predominant" because the highest transition probability for k_i is the reflexive transition as $d(k_i,k_i) = 1$. The normalisation factor $\delta(k_i)$ is computed only once for each keyterm k_i of the knowledge model, and the following equation holds: $\Sigma_{x \in K} Cty(k_i,x) = 1$.

3.3. A Small Example for Simple AR

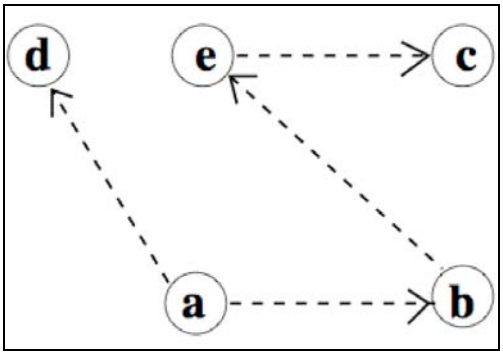

Figure 1. The knowledge model K.

Let Figure 1 be an example of a knowledge model, where an edge between k_i and k_j vertices is interpreted as the specialisation relation $k_i \sqsubseteq k_j$. Based on this model, we may compute the conformity related to each transition as shown in Table 4. Next, we provide details for the computation of the conformity measure for two examples: firstly between a and c where there exists a path in the model, and secondly between c and d, where a path is missing in (K, $\sqsubseteq$).

$$\begin{aligned} Cty(a, c) &= [d(a,c) \times \Sigma_{x\in\{a,b,c,d,e\}} 1/d(a,x)]^{-1} \\ &= [d(a,c) \times (1/d(a,a) + 1/d(a,b) + 1/d(a,c) + 1/d(a,d) + 1/d(a,e))]^{-1} \\ &= [3 \times (1 + 1 + 1/3 + 1 + 1/2)]^{-1} = 2/23 = 0.09 \end{aligned}$$

$$\begin{aligned} Cty(c, d) &= [d(c,d) \times \Sigma_{x\in\{a,b,c,d,e\}} 1/d(c,x)]^{-1} \\ &= [d(c,d) \times (1/d(c,a) + 1/d(c,b) + 1/d(c,c) + 1/d(c,d) + 1/d(c,e))]^{-1} \\ &= [11 \times (1/11 + 1/11 + 1 + 1/11 + 1/11)]^{-1} = 1/15 = 0.07 \end{aligned}$$

Table4.The conformity scores with the model (K, $\sqsubseteq$) of Figure 1

$\rightarrow$	a	b	c	d	e	Σ
a	0.26	0.26	0.09	0.26	0.13	1
b	0.03	0.37	0.19	0.03	0.37	1
c	0.07	0.07	0.73	0.07	0.07	1
d	0.07	0.07	0.07	0.73	0.07	1
e	0.04	0.04	0.44	0.04	0.44	1

Once the computation of the Table 4 is completed, the conformity for each simple rule $k_i \rightarrow k_j$ is given by looking up to the corresponding row i and column j of this table. From the previous example given in Table 2: r_1, r_{19} and r_{11}, r_{20} are two pairs of symmetrical simple rules. Hence, the Table 4 gives their conformity:

(r_{19}) : e→b with $Cty(r_{19}) = 0.04$	(r_{20}) : c→a with $Cty(r_{20}) = 0.07$
(r_1) : b→e with $Cty(r_1) = 0.37$	(r_{11}) : a→c with $Cty(r_{11}) = 0.09$

According to the conformity measure – the classification of the rules is presented in the increasing order – the interesting rules have the lowest values in conformity with (K, $\sqsubseteq$). For the four previous simple rules, the classification is established as follows: $\{r_{19}, r_{20}, r_{11}, r_1\}$.

The rule r_{11} (in 3rd position in the classification), is already known in (K, $\sqsubseteq$): its conformity is low because the distance between the two vertices (a and c) is the longest one in (K, $\sqsubseteq$). The first two rules r_{19} and r_{20} possibly could enrich the model under the supervision of the analyst. It should be noticed that these four rules are classified at very different ranks depending on the statistical measures. Likely because we use an extra knowledge source (K, $\sqsubseteq$), along with the textual database used for the classification in Table 3.

If an analyst studies the rules sequentially, following any statistical measure, he may be overwhelmed by rules which reflect knowledge already known in (K, $\sqsubseteq$). Moreover, a major knowledge loss occurs when a number of extracted rules containing new pieces of interesting knowledge are classified at the bottom following the statistical classification lists. On the contrary, the classification given by the conformity measure may draw the attention of the analyst on the possible enrichment of the current domain model (K, $\sqsubseteq$) with interesting extensions and modifications.

3.4. Conformity for Complex Rules

The complex rules have their left and/or right parts composed of more than one keyterm. Three different kinds of rules may be distinguished. The first is called a 1—m rule: $k_1 \rightarrow k_2 \wedge ... \wedge k_{m+1}$ with $m \geq 2$, and it is composed of one keyterm on the left part and its right part has at least two keyterms. The second is called a n—1 rule: $k_1 \wedge ... \wedge k_n \rightarrow k_{n+1}$ with $n \geq 2$ that has its left part composed of at least two keyterms and a right part with a single keyterm. Finally, an n—m rule: $k_1 \wedge ... \wedge k_n \rightarrow k_{n+1} \wedge ... \wedge k_{n+m}$ where both $(n, m) \geq 2$. We generalize the conformity measure for complex rules by examining its definition for the three kinds (respectively, 1—m, n—1, and n—m) AR.

1—m rules. Let us consider the example of a 1—2 rule: $R_1 : x \rightarrow y \wedge z$. Following predicate logic, R1 can be rewritten in: $\neg x \vee (y \wedge z) = (\neg x \vee y) \wedge (\neg x \vee z)$. This rule can be normalised in a clausal form and decomposed into a conjunction of simple rules: $R_1 = (x \rightarrow y) \wedge (x \rightarrow z)$. Accordingly, the rule R1 is in conformity with $(K, \sqsubseteq)$ if each simple rule of the decomposition is in conformity with $(K, \sqsubseteq)$. The conformity for R1 is then defined by:

$$Cty(R_1 : x \rightarrow y \wedge z) = \min(Cty(x \rightarrow y), Cty(x \rightarrow z))$$

The conformity measure range stands in [0,1[. The min function ensures that if at least one simple rule has a low conformity measure, then the complex rule has also a low conformity measure, i.e., the rule may contain some new information for updating $(K, \sqsubseteq)$. Conversely, if all the simple rules have a high conformity measures, i.e., if they all are conform to the model $(K, \sqsubseteq)$, then R_1 is also considered to be conform to $(K, \sqsubseteq)$.

n—1 rules. Let us consider the example of the 2—1 rule R_2: $x \wedge y \rightarrow z$. Following predicate logic, R_2 can be rewritten in: $\neg(x \wedge y) \vee z = (\neg x \vee \neg y) \vee z = (\neg x \vee y) \vee (\neg y \vee z)$. This rule can be decomposed into a disjunction of two simple rules: $R_2 = (x \rightarrow z) \vee (y \rightarrow z)$. Thus, the rule R_2 is in conformity with $(K, \sqsubseteq)$ if one of the simple rules of the decomposition is in conformity with $(K, \sqsubseteq)$. The conformity for R_2 is then defined by: $Cty(R_2 : x \wedge y \rightarrow z) = \max(Cty(x \rightarrow z), Cty(y \rightarrow z))$. The max function ensures that if at least one simple rule has a high conformity measure, then the complex rule has also a high conformity measure, i.e., $(K, \sqsubseteq)$ already contains the information carried out by R_2. Conversely, if all the simple rules have a low conformity measure, i.e., if there is no simple rule that is conform to the model $(K, \sqsubseteq)$, then R_2 is also considered as being not conform to $(K, \sqsubseteq)$.

n—m rules. Following the same two ideas, a n—m rule is considered as a conjunction of disjunction of simple rules. The 3—2 rule $R_3 : x \wedge y \wedge z \rightarrow v \wedge w$ can be decomposed into $[(x \rightarrow v) \vee (y \rightarrow v) \vee (z \rightarrow v)] \wedge [(x \rightarrow w) \vee (y \rightarrow w) \vee (z \rightarrow w)]$. Hence, the conformity for R_3 is defined by: $\min(\max(Cty(x \rightarrow v), Cty(y \rightarrow v), Cty(z \rightarrow v)), \max(Cty(x \rightarrow w), Cty(y \rightarrow w), Cty(z \rightarrow w)))$ and can be generalized for all simple and complex rules R into:

$$Cty(R : x_1 \wedge ... \wedge x_n \rightarrow y_1 \wedge ... \wedge y_m) = \min_{j=1}^{m}(\max_{i=1}^{n}(Cty(x_i, x_j)))$$

In doing so, we have to mention that the combination of min and max in the conformity measure for complex rules may lead to loose the fact that some keyterms for R, among all others, are related in $(K, \sqsubseteq)$.

Since other relations are absent in (K, ⊑), R should be presented to the analyst. This case is illustrated by the following rule r_{12}:

$Cty(b \wedge c \rightarrow a \wedge e) = \min(\max(Cty(b, a), Cty(c, a)), \max(Cty(b, e), Cty(c, e)))$
$= \min((\max(0.03, 0.07), \max(\mathbf{0.37}, 0.07)))$
$= \min(0.07, \mathbf{0.37}) = 0.07$

3.5. A Small Example for Complex AR

Given (K, ⊑) in Figure 1, the Table 5 shows the classification of the 20 valid AR extracted – 16 complex and 4 simple – in an increasing order according to their conformity with (K, ⊑). We notice that the conformity classification for complex rules is, as we expected, different from the classification with the statistical measures given in Table 3. The difference is due to the use of an extra knowledge source (K, ⊑) for the former classification, rather than the text collection only as for the latter classification. The next section gives the main results of a qualitative analysis on real-word corpus. We follow the same principle as used for the simple example: by comparing conformity *versus* statistical measure classifications[1], and by considering the analyst's perspective on the appropriate knowledge units carried by the rules.

Table5.Conformity of the 20 AR in Table 2 with the model (K, ⊑) depicted in Figure 1

id	Rule	Conformity	id	Rule	Conformity
r_6	$b \rightarrow a \wedge c \wedge e$	0.03	r_{18}	$c \wedge e \rightarrow a \wedge b$	0.07
r_7	$e \rightarrow b \wedge c$	0.04	r_{20}	$c \rightarrow a$	0.07
r_{10}	$e \rightarrow a \wedge b \wedge c$	0.04	r_9	$a \rightarrow c \wedge d$	0.09
r_{19}	$e \rightarrow b$	0.04	r_4	$a \rightarrow b \wedge c \wedge e$	0.09
r_{13}	$d \rightarrow a \wedge c$	0.07	r_{11}	$a \rightarrow c$	0.09
r_{14}	$c \rightarrow b \wedge e$	0.07	r_2	$b \rightarrow c \wedge e$	0.19
r_{15}	$c \rightarrow a \wedge d$	0.07	r_3	$a \wedge b \rightarrow c \wedge e$	0.19
r_{16}	$c \rightarrow a \wedge b \wedge e$	0.07	r_8	$a \wedge e \rightarrow b \wedge c$	0.26
r_{17}	$c \wedge e \rightarrow b$	0.07	r_5	$b \wedge c \rightarrow e$	0.37
$\mathbf{r_{12}}$	$b \wedge c \rightarrow a \wedge e$	0.07	r_1	$b \rightarrow e$	0.37

4. APPLICATION ON MOLECULAR BIOLOGY CORPUS

4.1. Description of the Experiment

On the one hand, there is a corpus of 1361 scientific paper abstracts holding on molecular biology[2] of about 240,000 words (1.6 M-Bytes). The theme of the texts is the phenomenon of gene mutation causing a bacterial resistance to antibiotics. The interpretation results from this specific domain needs a high degree of human expertise. On the other hand, there is a domain ontology – a set of semantically related concepts – used as a knowledge model (K, ⊑). The concepts of the ontology are the correct keyterms of

[1] The statistical measure classification is detailed in [Cherfi *et al.*, 2006], where an algorithm is proposed and an evaluation is carried out by an analyst –expert in molecular biology.

[2] The corpus is collected from the Pascal-BioMed documentary database of the French institute for scientific and technical information (INIST)

the domain and constitute the pieces of information we mine in the texts. Moreover, we assume that cooccurrence of the keyterms in a text reflects a semantic link between keyterms [Anick & Pustejovsky, 1990]. We used UMLS [UMLS, 2000] restricted to the keyterms of the domain and all their parent keyterms represented by the specialisation relation (IsA). A keyterm is a noun phrase in the domain ontology which can be associated to a concept, and thus, it ensures the transition from the linguistic to the knowledge level. In this way, the corpus has been indexed with 14,374 keyterms, including 632 different keyterms. The minimal support σ_s for the AR extraction is set to 0.7% – occurring, at least in 10 texts – and the minimal confidence σ_c is set to 80%. We obtain 347 valid AR, including 128 exact rules. From the set of 347 rules, we kept 333 AR which do not deal with ambiguities in the keyterm meaning – two or more concept identifiers (CUI) in the UMLS for the same keyterm. Thus, we discarded 14 AR, and there are 510 different concepts remaining (from 632 original ones). When the 510 concepts are augmented with their IsA-parents, K is composed of 1,640 vertices (concepts) and 4178 edges ($\sqsubseteq$ relations). Among them, concepts appear 364 times in the 333 AR. There are 53 concepts in common with K (i.e., 56%), whereas 41 concepts are absent in K (i.e., 44%) out of the 94 different concepts from the AR set. There is a total number of 2,689,600 transitions probabilities computed from the 510 keyterms in K. The number of transition probabilities stored for the calculation in the 333 AR is: 419,906. The conformity computation operates 739 comparisons (min or max) for the probability transitions, yielding a total number of 831 values – with $108 \neq 0$ (i.e., 13%) and 21 different transitions, including $Cty = 0$. Finally, the conformity value range is [0, 0.231] with 18 different measure values and 75 out of 333 rules have their $Cty > 0$. We have to notice that the conformity measure is set to 0 for keyterms that does not appear in the $(K, \sqsubseteq)$ rather than a minimal probability as stated in section 3.2, because the automatic computation of the probability transitions for $(K, \sqsubseteq)$ is done once and regardless of the corpus. Finally, there are four classes of AR in the 333 set: 45 (1—1) simple rules (i.e., 13.51%), 5 (1—n) complex rules (i.e., 1.5%), 250 (n—1) complex rules (i.e., 75.08%), and 33 (n—m) complex rules (i.e., 9.9%). Table 6 summarizes these results.

Table6.Results on the model (K, ⊑) and the rule set extracted from our corpus

333 AR set	# concepts	# different concepts		
	364	94		
$(K, \sqsubseteq)$ model	# concepts	# concepts (Is-A augmented)		
	510	1640		
Transition probability	# values	# non-zero values		
	831	108 (13%)		
AR class	1—1	1—n	n—1	n—m
	45	5	250	33

4.2. Quality Analysis of AR rankings in the KBTM Process

The analysis is conducted as follows: For each rule in the four AR classes (1—n, 1—n, etc.), we compare its conformity measure, its statistical measures and whether or not it belongs to three groups based on the analyst's expertise: (i) interesting rules, (ii) relating known meronyms (especially hypernyms) and synonyms, and (iii) useless rules. Thanks to the conformity measure, we focus on a subset of 258 rules (*i.e.*, 77.5%) over the 333 rule set that are not conform to $(K, \sqsubseteq)$ – as they relate keyterms that either are absent in K or isolated concepts following the relation $\sqsubseteq$. This gives a significant improving rate of 22.5%

of extracted AR that are candidate to be discarded from the rule set. The discarded rules may be examined by their own in a further analysis (see summary in Table 7).

Table7.Results of the subset presented to the domain expert for qualitative analysis

AR category	# AR	Percentage (%)
interesting (Cty=0)	258	77.5
useless (Cty>0)	75	22.5
Total	333	

In the following, and without exhaustiveness, we report the outcome through some examples: Firstly, we focus on two close n–1 rules interesting according to the analyst: one is conform and the other is not conform with regards to (K, ⊑). Next, we show and comment one simple AR belonging to the analyst's class: relating known keyterms. We end with an example of a useless AR according to the analyst. Some rules are identified as interesting by the analyst. For example, the mutation of the parC gene is interesting to comment in the following two (2—1) rules:

Rule Number: 221
"gyra gene" ∧ "substitution" → "quinolone"
Interest: "13.610" Conviction: "4.706" Dependency: "0.741" Novelty: "0.008"
Satisfaction:"0.788" Conformity: "0"
Rule Number: 218
"gyra gene" ∧ "sparfloxacin" → "ciprofloxacin"
Interest: "1.073" Conviction: "6.003" Dependency: "0.770" Novelty: "0.007"
Satisfaction: "0.833" Conformity: "0.000215"

The rule #218, with Cty > 0, draws the resistance mechanism for two antibiotics sparfloxacin and ciprofloxacin that are subsumed (⊑) by the concept of quinolone (a family of antibiotics) in K. Moreover, the rule #221 is more precise by pointing out the specific resistance mechanism (namely substitution). We notice that the major good statistical measures for these rules are: conviction and satisfaction. Nevertheless, both measures give the reverse classification compared to the conformity and the analyst comments below. Some simple AR relate synonyms or hypernyms keyterms. They belong to the group: relating known keyterms according to the analyst. This group of rules shows that authors of the texts describe the same concept with different keyterms, and the text mining process reveals such usage.

Rule Number: 183:
"epidemic strain" −→ "outbreak"
Interest: "17.449" Conviction: "undefined" Dependency: "0.943" Novelty: "0.011"
Satisfaction: "1.000" Conformity: "0"

The statistical measure that gives a good quality for rule #183 is the dependency (which is used as the 3rd quality measure to check following the algorithm given in [Cherfi *et al.*, 2006]). The interest measure classes this rule in the middle of the corresponding list. Conversely, the conformity is 0, which gives it a chance to be analysed and update (K, ⊑) with two missing relations epidemic strain ⊑ outbreak and outbreak ⊑ epidemic strain.

Finally, the rules #268 and #269 are examples which are considered as wrong, hence useless for the analysis. It is due to the fact that keyterms: mycobacterium and tuberculosis are not significant in the molecular biology domain; however, these keyterms are extracted as keyterm index and are present as concepts in the general UMLS. The correct concept, in this context, would be the keyterm mycobacterium tuberculosis (see in [Cherfi *et al.*, 2006]).

Rule Number: 268
"mutation" $\wedge$ "mycobacterium tuberculosis" $\rightarrow$ "tuberculosis"
Interest: "14.956" Conviction: "undefined" Dependency: "0.933" Novelty: "0.006"
Satisfaction: "1.000" Conformity: "0.000178"
Rule Number: 269
"mutation" $\wedge$ "mycobacterium" $\rightarrow$ "tuberculosis"
Interest: "12.463" Conviction: "5.599" Dependency: "0.766" Novelty: "0.010"
Satisfaction: "0.821" Conformity: "0.00017809"

The rules #268 and #269 have the same non zero conformity, and have also good statistical quality measures. Hence, they will be presented to the analyst. Using the KBTM process, and without knowledge loss, we can discard the rules #268 and #269 from the rule set presented to the analyst because they are useless by introducing the artefacts mycobacterium and tuberculosis which are irrelevant in the context of molecular biology.

5. DISCUSSION

Among studies that intend to handle the large set of AR extracted with statistical quality measures, [Kuntz *et al.*, 2000] is similar to the work presented in section 2.2. This methodology is of great interest to highlight rule properties such as resistance to noise in the data set, or to establish whether a rule is extracted randomly or not (i.e., by chance). However, the limits of these measures come from the fact that they do not consider any knowledge model.

The background knowledge is used during the data mining process in [Jaroszewicz & Simovici, 2004] with a Bayesian Network [Pearl, 1988] to filter interesting frequent itemsets. A Bayesian network is similar to the knowledge model $(K, \sqsubseteq)$ described in this chapter; except that each vertex (*i.e.*, relation) is associated with a weight defined by the relation conditional probability (*e.g.*, for the specialisation $\sqsubseteq$) *wrt.* to the concept parent(s) in the Bayesian network. The distribution probabilities over the relations are set up, *a priori*, by expert's judgments. The authors propose an algorithm to compute the marginal distributions of the itemset (*e.g.*, corresponding to the keyterm sets when dealing with text applications) over the Bayesian network. Hence, the itemset marginal distributions are inferred from the Bayesian network structure. An itemset is interesting if its support in the corpus (*i.e.*, real support of appearing in the texts) deviates, with a given threshold, from the support inferred from the Bayesian network (*i.e.*, its conditional probability to occur in the knowledge domain). A sampling-based approach algorithm for fast discovery of the interesting itemsets (called unexpected patterns) is given in [Jaroszewicz & Scheffer, 2005].

This methodology is extended in [Faure *et al.*, 2006] to drive both the AR extraction and the Bayesian network's weight updates. Hence, iteratively, the interesting AR identified in this way are candidates to update the Bayesian network. The similarities with the approach presented in this chapter are high.

However, when [Faure *et al.*, 2006] deal with probabilistic reasoning and analyst's judgments on the structure of the Bayesian Network, we rather stick to more formal knowledge conveyed by an ontological (i.e., consensual) domain knowledge model. However, the approach in [Faure *et al.*, 2006] could be complementary to the KBTM approach presented in this chapter. Further studies can be conducted to study the AR rankings given by both approaches for a given domain corpus *wrt.* to, respectively, a knowledge model, and a Bayesian network.

Another interesting work for the post-mining of association rules involving user interaction as backgroung knowledge is [Sahar, 1999; Liu *et al.*, 2003]. Here, the user is asked to interact with the system in order to evaluate the quality of the rules. [Sahar, 1999] assumes the following hypothesis: if a simple rule $k_1 \rightarrow k_2$ is of low interest for the user, then all related complex rules – related rules are defined as rules containing k_1 in their body and k_2 in their head– are also considered as of low interest. The user does not have to study them and the number of rules to study is substantially reduced. The user is asked to classify simple rules in one of the four categories: (1) true but uninteresting, (2) false and interesting, (3) false and uninteresting, (4) true and interesting. If a simple rule is classified in class (1) or (3), then the rule itself and its complex related rules may be deleted from the set of rules. This work has some other interesting characteristics: (i) An appropriate algorithm has been developed to select the simple rules to be given first to the user. The selected rules are the ones connected to a large number of complex rules. In this way, the number of rules to study decreases more rapidly than a random choice. (ii) The approach takes into account the direction of the rule: the deletion by the user of the rule $k_1 \rightarrow k_2$ has no effect on the rule $k_2 \rightarrow k_1$. (iii) [Sahar, 1999] does not use a knowledge model but the subjective judgement of the user which may be seen as an informal knowledge model. (iv) Finally, the major difference between our approach and [Sahar, 1999] concerns the interpretation of complex rules. The assumption adopted in [Sahar, 1999] is that any complex rule, according to our interpretation, could be turned to a conjunction of simple rules. However, we have shown that such decomposition, in clausal form, is misleading: $1 - m$ rules can be rewritten into a conjunction of simple rules; whereas $n - 1$ rules are rewritten into a disjunction of simple rules.

[Basu *et al.*, 2001] proposes another approach and uses WORDNET lexical network to evaluate the quality of the rule where keyterms are, actually, words. The quality score of a simple rule $word_1 \rightarrow word_2$ is given by the semantic distance between $word_1$ and $word_2$ in the lexical network. The network is a weighted graph, and each semantic relation (syno/antonymy, hyper/hyponymy) has its own weight. The distance between two words is the lower weight path in the graph. For any complex rule, the quality score is the mean of the distance for each pair $(word_i , word_j)$ where $word_i$ is in the body of the rule and $word_j$ is in its head. Here, as in [Sahar, 1999], the definition of the score for complex rules is logically false. The advantage in [Basu *et al.*, 2001] is the ability to deal with several semantic relations. However, the different properties of these relations cannot be formally expressed using a weighted graph and some assumptions are made such as: $weight(synonymy) > weight(hypernymy)$, etc. This method, based on a network of lexical entities, could be adapted to a formal knowledge model. However, it cannot be used to update a knowledge model: the weighting system and the mean calculation of the score for complex rules make impossible the association of a rule with a knowledge model as we did in Table 5.

6. CONCLUSION AND FUTURE WORK

In this chapter, we have proposed two methods for classifying association rules extracted within a KBTM process: the first one is based on statistical measures, and the second one is based on conformity with a knowledge model. Our present research study sets a knowledge-based text mining (KBTM) process driven by a knowledge model of the domain. Association rules that do not correspond to known relations of specialisation in the knowledge model are identified thanks to the conformity measure. The behaviour of the conformity measure is in agreement with the KBTM process. The conformity measure allows us both the enrichment of the knowledge model, and the TM process efficiency enhancement. An experiment on real-world textual corpus gives a significant improving rate and shows the benefits of the proposed approach to an analyst of the domain.

Furthermore, the conformity measure proposed in this first study can be extended to a number of promising directions in order to assess its effectiveness in different knowledge domains and contexts. Firstly, it could be interesting to take into account in the knowledge model of molecular biology domain other relations such as: causality (by considering rules involving instances of antibiotics → bacteria), temporal (the study of gene parC) mutation is anterior to gyrA study, how this relation has an impact on the resistance mechanism to antibiotics). In doing so, we will be able to have a deeper understanding of the texts and suggest an accurate modification of the knowledge model itself within the KBTM process.

REFERENCES

[Agrawal *et al.*, 1996] Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., & Verkamo, A. I. (1996). Fast Discovery of Association Rules. In U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, & R. Uthurusamy (Ed.), *Advances in Knowledge Discovery and Data Mining* (pp. 307–328). Menlo Park, CA: AAAI Press / MIT Press.

[Anick & Pustejovsky, 1990] Anick, P., & Pustejovsky, J. (1990). An Application of lexical Semantics to Knowledge Acquisition from Corpora. *30th International Conf. on Computational Linguistics (COLING'90)*: Vol. 3 (pp. 7–12). Helsinki, Finland.

[Bastide *et al.*, 2000] Bastide, Y., Taouil, R., Pasquier, N., Stumme, G., & Lakhal, L. (2000). Mining frequent patterns with counting inference. *ACM SIGKDD Exploration Journal*, 2(2): 66–75.

[Basu *et al.*, 2001] Basu, S., Mooney, R.J., Pasupuleti, K.V., and Ghosh J. (2001). Evaluating the Novelty of Text-Mined Rules using Lexical Knowledge. *7th ACM SIGKDD International Conference on Knowledge Discovery in Databases* (pp. 233–238). San Francisco, CA: ACM Press.

[Cherfi *et al.*, 2006] Cherfi, H., Napoli, A., & Toussaint, Y. (2006). Towards a text mining methodology using frequent itemsets and association rules. *Soft Computing Journal - A Fusion of Foundations, Methodologies and Applications,* 10(5):431–441. Special Issue on "Recent Advances in Knowledge and Discovery". Springer-Verlag.

[Collins & Loftus, 1975] Collins A. & Loftus E. (1975). A spreading-activation of semantic processing. *Psychological Review*, 82(6):407–428.

[Faure *et al.*, 2006] Faure, C., Delprat, D., Boulicaut, JF., & Mille, A. (2006). Iterative Bayesian Network Implementation by using Annotated Association Rules. *15th Int'l Conf. on Knowledge Engineering and Knowledge Management – Managing Knowledge in a World of Networks, Vol. 4248 of Lecture Notes in Artificial Intelligence – LNAI* (pp. 326–333). Prague, Czech Republic: Springer-Verlag.

[Fayyad *et al.*, 1996] Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., & Uthurusamy, R. (1996). *Advances in Knowledge Discovery and Data Mining*. Menlo Park, CA: AAAI Press / MIT Press.

[Jacquemin, 2000] Jacquemin, C., (1994). FASTR: A unification-based front-end to automatic indexing. *Information multimedia, information retrieval systems and management* (pp.34–47). New-York, NY: Rockfeller University.

[Jaroszewicz & Scheffer, 2005] Jaroszewicz, S. & Scheffer, T. (2005). Fast Discovery of Unexpected Patterns in Data, Relative to a Bayesian Network. *ACM SIGKDD Conference on Knowledge Discovery in Databases* (pp. 118–127). Chicago, IL: ACM Press.

[Jaroszewicz & Simovici, 2004] Jaroszewicz, S. & Simovici, D.A. (2004) Interestingness of Frequent Itemsets using Bayesian networks as Background Knowledge. *ACM SIGKDD Conference on Knowledge Discovery in Databases* (pp. 178–186). Seattle, WA: ACM Press.

[Kuntz *et al.*, 2000] Kuntz, P., Guillet, F., Lehn, R., & Briand, H. (2000). A User-Driven Process for Mining Association Rules. D. Zighed, H. Komorowski, & J. Zytkow (Ed.). *4th Eur. Conf. on Principles of Data Mining and Knowledge Discovery (PKDD'00), Vol. 1910 of Lecture Notes in Computer Science – LNCS* (pp. 483–489), Lyon, France: Springer-Verlag.

[Lavrac *et al.*, 1999] Lavrac, N., Flach, P., & Zupan, B. (1999). Rule Evaluation Measures: A Unifying View. *9th Int'l Workshop on Inductive Logic Programming (ILP'99). Co-located with ICML'9., Vol. 1634 of Lecture Notes in Artificial Intelligence – LNAI (*pp. 174–185). Bled, Slovenia: Springer-Verlag, Heidelberg.

[Liu *et al.*, 2003] Liu, B., Ma, Y., Wong, C., & Yu, P. (2003). Scoring the Data Using Association Rules. *Applied Intelligence*, 18(2): 119–135.

[Pasquier *et al.*, 1999] Pasquier, N., Bastide, Y., Taouil, R., & Lakhal, L. (1999). Pruning closed itemset lattices for association rules. *International Journal of Information Systems*, 24(1): 25–46.

[Paroubek, 2007] Paroubek, P. (2007). Evaluating Part-Of-Speech Tagging and Parsing – On the Evaluation of Automatic Parsing of Natural Language (Chapter 4). In L. Dybkaer, H. Hemsen, & W. Minker (Ed.), *Chapter 4 of Evaluation of Text and Speech Systems* (pp. 99–124). Springer.

[Pearl, 1988] Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. San Fransisco, CA: Morgan Kaufmann.

[Sahar, 1999] Sahar, S. (1999). Interestingness via What is Not Interesting. S. Chaudhuri, & D. Madigan, (Ed.), *5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'99)*. (pp. 332–336). San Diego, CA:. ACM Press.

[UMLS, 2000] UMLS (2000). *The Unified Medical Language System*., (11th edition): National Library of Medicine.