

# ON THE EVALUATION OF MINED FREQUENT SEQUENCES

## *An Evidence Theory-based Method*

Francisco Guil

*Computer Science School, University of Almería, Spain*

Francisco Palacios

*University Hospital of Getafe, Spain*

Manuel Campos, Roque Marín

*Computer Science Faculty, University of Murcia, Spain*

**Keywords:** Temporal data mining, Theory of evidence, Entropy, Specificity.

**Abstract:** Frequent sequences (or temporal associations) mining is a very important topic within the temporal data mining area. Syntactic simplicity, combined with the dual characteristics (descriptive and predictive) of the mined temporal patterns, allow the extraction of useful knowledge from dynamic domains, which are time-varying in nature. Some of the most representative algorithms for mining sequential patterns or frequent associations are Apriori-like algorithms and, therefore, they cannot handle numeric attributes or items. This peculiarity makes it necessary to add a new process in the data preparation step, the discretization process. An important fact is that, depending on the discretization technique used, the number and type of discovered temporal patterns change dramatically. In this paper, we propose a method based on the *Shafer's Theory of Evidence* that uses two information measures proposed by *Yager* for the quality evaluation of the extracted sets of temporal patterns. From a practical point of view, the main goal is to select, for a given dataset, the best discretization technique that leads to the discovery of useful knowledge. Nevertheless the underlying idea is to propose a formal method for assessing the mined patterns, seen as a belief structure, in terms of certainty in the information that represents. In this work, we also present a practical example, describing an application of this proposal in the Intensive Care Burn Unit domain.

## 1 INTRODUCTION

Temporal data mining can be defined, in general terms, as the activity of looking for interesting correlations or patterns in large sets of data accumulated for other purposes. It has the capability of mining activity, inferring associations of contextual and temporal proximity, some of which may also indicate a cause-effect association. This important sort of knowledge can be overlooked when the temporal dimension is ignored or treated as a simple numeric attribute. There is a lot of work related to this area, including those belonging to the *Apriori*-like algorithms, like sequential pattern mining or temporal association rules mining, amongst other. A common feature of the *Apriori*-like algorithms is that they can not handle databases with continuous attributes, requiring the application of discretization techniques in the data pre-processing step. Basically, the discretization tech-

niques consist of the partition of the numerical domain into a set of intervals, treating each interval as a category. There are automatic discretization techniques which find the partition that optimizes a given evaluation function, and techniques that use expert knowledge to divide the domain into several intervals. In general, there is a large number of discretization techniques, and their efficiency is often a bottleneck in the knowledge discovery process. The problem here is to select a good discretization method that, in an efficient way, generates a discretized dataset from which useful knowledge can be obtained. If we are dealing with global data mining techniques, the accuracy of the model can be a good proposal to select the best discretization technique. But, if we are dealing with local methods, it is so difficult to assess the quality of the discovered knowledge. The usually huge number of mined patterns, a number which depends both on the user-defined parameters

of the algorithm and the used discretization method, makes it virtually impossible to objectively evaluate the outcome. It becomes necessary, therefore, a formal method for evaluating mined (temporal) associations, which resulted in a value that indicates the quality of the discovered knowledge. In this paper, we propose to use a method to assess the quality of mined frequent sequences based on the combined use of two uncertainty measures, the entropy and specificity, both defined in (Yager, 1981) in the setting of the Theory of Evidence. Once the set of associations is prepared (normalizing the associated frequency distribution), we treat it as a body of evidence. From the Theory of Evidence point of view, data mining is seen as an evidence-based agent. But, from the data mining point of view, the treatment of sets of frequent associations as bodies of evidence allowing the use of the information measures to quantify the quality of the discovered knowledge. Specifically, we propose a quality index based on the distance of the pair formed by the entropy and non-specificity measures with regard to the situation of total certainty. For the empirical evaluation of the proposed method, we have used a dataset belonging to the medical domain, in particular, to the Intensive Care Burn Unit domain, specialized in the treatment of patients with severe burns.

The rest of this paper is organized as follows. Section 2 introduces the notation and basic definitions necessary to define the problem. Section 3 introduces the theoretical foundations of the proposed method for evaluating sets of frequent event-based sequences. Section 4 describes and empirical evaluation with a real dataset belonging to the medical domain. Conclusions and future works as finally drawn in Section 5.

## 2 THE PROBLEM DEFINITION

Frequent sequence mining is an extension of the *Apriori* algorithm (Agrawal et al., 1993) and, therefore, the mining process is iterative in nature and performs a levelwise search. It is based on the *downward closure property*, which states that every subset of a frequent sequence is also frequent (and also, it is true that every superset of a non-frequent sequence is non-frequent). Starting with the dataset, the first step focuses on the extraction of all the frequent events (sequences of length 1), and then the process continues generating the set of patterns of length  $k$  from the set of frequent patterns of  $k - 1$  length. Next, we will introduce the notation and basic definitions for specifying in detail the main goal of the proposal.

**Definition 1 (Event).** An event ( $e$ ) is defined as the pair  $(te, t)$ , or simply  $te$ , where  $te$  is the type event and  $t$  is the time instant of its occurrence. The events are things that happen in the real world, and they usually represent the dynamic aspect of the world. In our case, an event is related to the fact that a certain type event occurs at a given point-based instant.

**Definition 2 (Event-based Sequence).**

A sequence ( $S$ ) is defined as an ordered set of events, that is,  $S = \{e_0, e_1, \dots, e_{k-1}\}$ , where  $\forall i < j, e_i < e_j$ . Obviously,  $|S| = k$ .

This sort of sequence is defined by a quantitative treatment of time. Nevertheless, the sequential patterns are qualitative temporal sequences although, with a simple transformation, they can be expressed as in the Definition 2.

**Example 1** Let  $S$  be a qualitative sequence (or a sequential pattern), such that  $S_{ql} = \{ab \rightarrow cd \rightarrow a\}$ . The  $\rightarrow$  symbol denotes the "after than" temporal relation. This sequence can be rewritten using a quantitative notation in the form  $S_{qn} = \{a_0, b_0, c_1, d_1, a_2\}$ . In general, for  $i > 0$  and  $j > i$ ,  $S_{ql}$  generates a family of equivalent quantitative sequences defined as  $S_{qn}^* = \{a_0, b_0, c_i, d_i, a_j\}$ .

Given a dataset  $D$ , the goal of temporal sequence mining is to determine all the frequent sequences that show the temporal regularities in it.

**Definition 3 (Temporal Data Mining Algorithm).**

In our paper, a temporal data mining algorithm, denoted as *TDM*, is an *Apriori*-like algorithm that extracts a set of frequent temporal associations (also called temporal patterns, sequences or even event-based sequences) characterized by a frequency distribution. Each frequency value indicates the number of objects that match the corresponding pattern.

The term frequent is related to the fact that every frequency value associated with a pattern will be always greater than or equal to a user-defined parameter called *minimum support* (denoted as  $ms$ ). Depending on the sort of algorithm, another user-defined parameter can be *maxspan*, which indicates the maximal temporal distance among events that comprise a sequence. Examples of *TDM* algorithms of our interest are mainly *GSP* (Srikant and Agrawal, 1996), *SPADE* (Zaki, 2001), and *TSET - Miner* (an improved sequential version of the intertransactional algorithm presented in (Guil et al., 2004)). The two first algorithms are designed for mining sequential patterns. The latter are designed for extracting a special type of pattern, called event-based sequences. However,

as shown above, the sequential pattern can be represented as an event-based sequence, placing all the three algorithms as viable alternatives in this paper.

**Definition 4 (Sequences Base).** Let  $\mathcal{BS}$  be the total number of frequent sequences extracted by TDM, that is,  $\mathcal{BS} = \{S_i\}$ , where  $S_i$  is a frequent sequence.  $\mathcal{BS}$  is characterized by a frequency distribution  $f = \{f_i\}$ , such that  $f_i = \text{freq}(S_i)$ . We propose to normalize the distribution  $f$ , obtaining  $\bar{f} = \{\bar{f}_i\}$ , such that  $\bar{f}_i = \frac{f_i}{\sum_i f_i}$ .

From the set of all frequent sequences, in this paper, we are interested only in the maximal ones. A maximal sequence is a frequent sequence that has no frequent super-sequences. In this way, the mining process is done in a more efficient way, obtaining an interesting sequences base that shows the temporal regularities presented in the dataset. Once the sequences base is extracted, the next step is its evaluation and interpretation. These tasks are designed with the aim to discover useful knowledge, and usually they are performed by the expert. However, this is a very hard process and rarely can be done due to the huge amount of frequent sequences that are often discovered. The problem is compounded if, instead of evaluating one sequences base, the expert must evaluate a set of sequences bases obtained by varying the parameters of the algorithms for preprocessing the datasets. In this case, the evaluation process must be carried out in parallel for each of the mined bases, thus having a much more complicated problem. If the expert must compare two (or more) sequences bases to determine which one is the better for obtaining useful knowledge, what method should he/she follow? Is better a base with more patterns, or else the base with the longest patterns? Does not affect the structure and frequency of the patterns to the quality of information that they represent? On an experimental basis, we can determine the best base obtaining, for each one, a classification model and studying its accuracy. But, several problems arises from this approach. On the one hand, there is no single method to obtain a classification model from a set of sequences. What is more, in the case of temporal classification models, its generation from temporal patterns is still a open (and very interesting) problem. On the other hand, and assuming that it is possible to obtain a suitable model from each base, this evaluation method is time-consuming and it must be done using a trial-error method of problem solving. So, it seems interesting to study a formal method of evaluation that indicates us the quality of the base without having to generate any later model, that is, a formal method that determines the degree of certainty of a sequences base from the structure and

frequency of each sequence in it. And, precisely, this is the main goal of the next section.

### 3 MEASURING THE CERTAINTY OF MINED SEQUENCES

Starting with a sequence base characterized by a normalized frequency distribution, our goal here is to introduce the information measures that will enable us to achieve its quality in terms of certainty in the evidence. In (Yager, 1981), the author introduces the concepts of entropy (from the probabilistic framework) and specificity (from the possibilistic framework), in the framework of Shafer's theory. Both information measures of uncertainty provide complementary measures of the quality of a body of evidence. With this proposal, Yager extend the *Theory of Evidence of Shafer* (also known as *Dempster-Shafer Theory*), developed for modeling complex systems.

In the next section, a brief summary of the Shafer's Theory of Evidence is introduced. Next, the section concludes with the proposed method, defining the measures and setting their goals.

#### 3.1 Shafer's Theory of Evidence

The Shafer's Theory of Evidence is based on a special fuzzy measure called *belief measure*. Beliefs can be assigned to propositions to express the uncertainty associated to them being discerned. Given a finite universal set  $\mathcal{U}$  (the *frame of discernment*), the beliefs are usually computed based on a density function  $m : 2^{\mathcal{U}} \rightarrow [0, 1]$ , called *basic probability assignment* (bpa):

$$m(\emptyset) = 0, \text{ and } \sum_{A \subseteq \mathcal{U}} m(A) = 1. \quad (1)$$

$m(A)$  represents the belief exactly committed to the set  $A$ . If  $m(A) > 0$ , then  $A$  is called a *focal element*. The set of focal elements constitute a core:

$$\mathcal{F} = \{A \subseteq \mathcal{U} : m(A) > 0\}. \quad (2)$$

The core and its associate bpa define a *body of evidence*, from where a belief function  $Bel : 2^{\mathcal{U}} \rightarrow [0, 1]$  is defined:

$$Bel(A) = \sum_{B|B \subseteq A} m(B). \quad (3)$$

From any given measure  $Bel$ , a dual measure,  $Pl : 2^{\mathcal{U}} \rightarrow [0, 1]$  can be defined:

$$Pl(A) = \sum_{B|B \cap A \neq \emptyset} m(B). \quad (4)$$

It can be verified (Shafer, 1976) that the functions  $Bel$  and  $Pl$  are, respectively a possibility (or necessity) measure if and only if the focal elements from

a nested or consonant set, that is, if it can be ordered in such a way that each set is contained within the next. In that case, the associated belief and plausibility measures possess the following properties. For all  $A, B \in 2^{\mathcal{U}}$ ,

$$Bel(A \cap B) = N(A \cap B) = \min[Bel(A), Bel(B)], \quad (5)$$

and

$$Pl(A \cup B) = \Pi(A \cup B) = \max[Pl(A), Pl(B)], \quad (6)$$

where  $N$  and  $\Pi$  are the necessity and possibility measures, respectively. If all the focal elements are singletons, that is,  $A = \omega$ , where  $\omega \in \mathcal{U}$ , then:

$$Bel(A) = Pr(A) = Pl(A) \quad (7)$$

where  $Pr(A)$  is the probability of  $A$ . In general terms,

$$Bel(A) \leq Pr(A) \leq Pl(A). \quad (8)$$

A significant aspect of Shafer's structure is the ability to represent in this common framework various different types of uncertainty, that is, probabilistic and possibilistic uncertainty.

### 3.2 Specificity and Entropy

The concept of information measures has been entirely renewed by investigations in the setting of modern theories of uncertainty like Fuzzy Sets, Possibility Theory, and Shafer's theory of evidence. An important aspect is that all of these theories, although they have a different approach than the probability theory, they are quite related. The new types of generalized information measures enable several facets of uncertainty to be discriminated, modeled, and measured. Let  $\mathcal{B}(\mathcal{U})$  be the set of normal bodies of evidence on  $\mathcal{U}$  ( $m(\emptyset) = 0$ ). An information measure will be any mapping  $f : \mathcal{B}(\mathcal{U}) \rightarrow [0, +\infty)$  (the non-negative real line) (Dubois and Prade, 1999).  $f$  supposedly depends on both the core ( $\mathcal{F}$ ) of the body of evidence to which it applies, and its associated bpa ( $m$ ).  $f$  pertains to some property of bodies of evidence and assesses the extent to which it is satisfied. In the paper (Dubois and Prade, 1999), the authors study three different measures in the general framework of evidence theory, corresponding to different properties of bodies of evidence: measures of imprecision, dissonance and confusion. However, in this paper, we only take into account two particular measures of imprecision and dissonance, the specificity and entropy (respectively), because, as Yager pointed, the combination of them is a good approach to measure the quality of a particular body of evidence. Initially, Yager introduced the concept of specificity (Yager, 1981), as an amount that estimates the precision of a

fuzzy set. Later, the author (Yager, 1983) and Dubois and Prade (Dubois and Prade, 1985), extend the concept to deal with bodies of evidence, defining the measure of specificity ( $S_p$ ) of a body of evidence as:

$$S_p(m) = \sum_{A \subseteq \mathcal{U}} \frac{m(A)}{|A|}. \quad (9)$$

$|A|$  denotes the number of elements of the set  $A$ , that is, the cardinality of  $A$ . It is easy to see that  $1 - S_p(m)$  is a measure of imprecision, related to the concept of non-specificity. We will denote this imprecision value as  $J_m$ , that is,  $J_m = 1 - S_p(m)$ . Also, in (Yager, 1983), the author extends the Shannon entropy to bodies of evidence, considering the following expression where  $\ln$  is the Naperian natural logarithm (it is possible to use  $\log_2$  instead of  $\ln$  as well):

$$E_m(m) = - \sum_{A \subseteq \mathcal{U}} m(A) \ln(Pl(A)). \quad (10)$$

$Pl(A)$  is the plausibility of  $A$ , and  $\ln(Pl(A))$  can be interpreted in terms of Shafer's weight of conflict. As the Shannon entropy,  $E_m(m)$  is a measure of discordance associated with the body of evidence.

In the ideal situation, where no uncertainty is presented in the body of evidence,  $E_m(m) = J_m(m) = 0$ . This is the key point of the proposed method for measuring the certainty associated with a belief structure.  $E_m$  provides a measure of the dissonance of the evidence, whereas  $J_m$  provides a measure of the dispersion of the belief. So, the lower the  $E_m$ , the more consistent the evidence and, the lower  $J_m$  (the higher  $S_p$ ), the less diverse. For certainty, we want low  $E_m$  and  $S_m$  measures. So, by using a combination of both measures, we can have a good indication of the quality of evidence. For measuring this indicator ( $Q(m)$ , denoting quality), we propose the inverse of the Euclidean distance between the pair  $(E_m, J_m)$  and the pair  $(0, 0)$ , that is:

$$Q(m) = \frac{1}{\sqrt{(E_m)^2 + (J_m)^2}}. \quad (11)$$

### 3.3 Evaluation of a Sequence Base

Let  $\mathcal{BS}$  be the sequences base, formed by a set of frequent sequences,  $\mathcal{BS} = \{S_i\}$ ,  $1 \leq i \leq |\mathcal{BS}|$ , mined from a dataset. And let  $\bar{f}$  be the normalized frequency distribution associated with  $\mathcal{BS}$ . A frequent sequence is an ordered set of events. Let  $\mathcal{E}$  be the set of all the frequent events present in the base. Making a syntactic correspondence with the basic elements of the Theory of Evidence, we obtain that the pair  $(\mathcal{BS}, m)$  is a body of evidence, where  $\mathcal{BS}$  is the core (formed by the set of focal elements  $S_i \subseteq \mathcal{E}$ ), and  $\bar{f}$  is the

basic probability assignment. In this case,  $\mathcal{E}$  is the frame of discernment. Once this correspondence is established, the evaluation of the set of mined patterns would be carried out by the method described in the previous section, that is, using the function  $Q(m)$  defined in Equation 11. In general terms, this measure adds additional information to the expert on the quality of the sequences base. But, in particular, in this paper, the  $Q(m)$  function (specifically  $Q(\bar{f})$ ) will be used to compare objectively three set of frequent patterns mined from a dataset belonging to a medical domain, which has been discretized using three different discretization methods. Let  $TDM$  be an Apriori-like temporal data mining algorithm, and let  $D$  be a dataset. A special feature of  $TDM$  is that it can not handle numerical attributes. Since  $D$  contains continuous attributes, very common in real datasets, it will be require a discretization method to obtain a dataset with only nominal attributes. Let  $d_1, d_2, d_3$  be three different discretization methods that generate three different datasets,  $D_1 = d_1(D)$ ,  $D_2 = d_2(D)$ , and  $D_3 = d_3(D)$ . The execution of the algorithm on each of the datasets ( $TDM(D_i)$ ) will results in three different sequences bases, denoted as  $\mathcal{B}S^{D_i}$ , each one characterized by a normalized frequency distribution  $\bar{f}_i$ . In order to compare the three discretization methods and determine which one provides information with less uncertainty, we propose the use of the  $Q(\bar{f}_i)$  function, such that the best method is the one that generates a base with the highest value of  $Q$ . For a more complete assessment, in the empirical evaluation we will use also three different values for the minimum support parameter of the  $TDM$  algorithm.

#### 4 EMPIRICAL EVALUATION

From a practical point of view, we have carried out an empirical evaluation using a preprocessed dataset that represents the evolution of 363 patients in an Intensive Care Burn Unit (*ICBU*) between 1992 and 2002. The original database stores, for each patient, a lot of clinical parameters such as age, presence of inhalation injury, the extent and depth of the burn, the necessity of early mechanical ventilation, and the patient status int its last day of stay in the Unit, among others. However, in the construction of the dataset, we only take into account the temporal parameters, which indicate the evolution of the patients during the resuscitation phase (first 2 days) and during the stabilization phase (3 following days). Incomings, diuresis, fluid balance, acid base balance (pH, bicarbonate, base excess) and other variables help to define objectives and to assess the evolution and treatment response. For

each of these temporal variables, we used three different discretization methods (called  $d_1$ ,  $d_2$ , and  $d_3$ , respectively). The first one ( $d_1$ ) is based on clinical criteria and uses the knowledge previously defined in the domain. A second discretization method ( $d_2$ ) can be based on the usual interpretation of mean value and standard deviation. In statistics, the mean is a central value around which the rest of the values are spread. When the distance of an element to the mean is greater than two standard deviations it should be carefully looked at because it may be a potential outlier. We have made a similar distinction defining normal values those in a interval of one standard deviation around the mean, slightly high or slightly low values those within two standard deviations around the mean, and high or low values for those whose distance to the mean is greater than two standard deviations. This discretization method is the one that generates the highest number of patterns, since it does not consider the domain values (that can be arbitrary in the dataset) but the values found in the data and therefore most of them should be found in the “normal” interval if they are distributed around the mean. In the last method ( $d_3$ ), we used an entropy-based information gain with respect the output variable. The information gain is equal to the total entropy for an attribute if for each of the attribute values a unique classification can be made for the output variable.

For each discretized dataset ( $D_i$ ), we have obtained a set of maximal sequences using a version of a temporal data mining algorithm designed for the extraction of frequent sequences from datasets with a time-stamped dimensional attribute. In the analysis of the data, different values for the parameters of the algorithm were set, in particular, maxspan to 5 days (resuscitation and stabilization), and minimum support to 20%, 30%, and 50%. In total, 6 sets of frequent sequences were extracted that, after the normalization process, resulted in the generation of 6 normalized sequences bases (denoted as  $\mathcal{B}S_{ms}^{D_i}$ , where  $D_i$  is the dataset obtained by the discretization technique  $d_i$ , and  $ms$  is the minimum support). In terms of the Theory of Evidence, each normalized sequences base corresponds to a body of evidence or belief structure, and is the input of the evaluation method that we propose in this paper. Table 1 shows the results of the proposed evaluation method. For each discretization technique, and for each minimum support value, we indicate in the table the entropy ( $E_m$ ) and the non-specificity based measure ( $J_m$ ), where  $J_m = 1 - S_m$ , and  $S_m$  the specificity like measure, both proposed by Yager as a quality indicator of a belief structure.

The  $Q$  parameter is the quality measure defined in the Equation 11, that is, the inverse of the classi-

Table 1:  $E_m$ ,  $J_m$ , and  $Q$  measures.

$\mathcal{B}S_{ms}^{D_i}$	Measures	20%	30%	50%
	$E_m$	0,4071	0,6294	1,0070
$\mathcal{B}S_{ms}^{D_1}$	$J_m$	0,7897	0,7311	0,6139
	$Q$	1.1255	1.0366	0.8479
	$E_m$	0,0314	0,1252	0,4822
$\mathcal{B}S_{ms}^{D_2}$	$J_m$	0,8699	0,8314	0,7318
	$Q$	1.1489	<b>1.1895</b>	1.1410
	$E_m$	0,1977	0,3774	0,7040
$\mathcal{B}S_{ms}^{D_3}$	$J_m$	0,8382	0,7851	0,6681
	$Q$	1.1612	1.1480	1.0303

cal Euclidean distance between the pair  $(E_m, J_m)$  and the vector  $(0,0)$ , which corresponds to the ideal situation where no uncertainty is presented in the body of evidence. In our case, the best solution, in terms of quality in the information presented in the patterns (maximal  $Q$ ), is the sequences base  $\mathcal{B}S_{30\%}^{D_2}$ , that is, the set of frequent maximal sequences obtained from the dataset discretized by the  $d_2$  method (the statistical version) and with the minimum support parameter set to 30%. Depending of the intervals extracted by the discretization methods, the relationship between minimum support and number of patterns varies significantly. This variation correlates with a variation in the measures of entropy and non-specificity so that a decrease in the minimum support value involves a less specific and less entropic set of patterns (see Table 1). On the one hand, we are interested in the extraction of sets of patterns with low entropy or, what is the same, with less uncertainty. So, if we take this measure independently, we have that the best sequence base is  $\mathcal{B}S_{20\%}^{D_2}$ , which corresponds with the dataset with a greater number of patterns. But, on the other hand, we are also interested in the extraction of more specific sets of patterns and, in this case, it corresponds to the  $\mathcal{B}S_{50\%}^{D_1}$  base, which has a smaller number of patterns. Therefore, it is necessary to make a compromise, taking into account the two measures together. So, taking into account the  $Q$  measure, the less entropic and more specific set of patterns is, precisely,  $\mathcal{B}S_{30\%}^{D_2}$ .

## 5 CONCLUSIONS AND FUTURE WORK

In this paper we have presented a formal method for evaluating the data mining results, that is, the set of frequent patterns that shows the regularities presented in a dataset. The method is based on the use of two

information measures proposed by Yager in the context of the *Shafer's Theory of Evidence*. The proposal involves the combined use of both measures, the entropy-like measure ( $E_m$ ), and the (non)specificity-like measure ( $J_m$ ), to quantify the quality level (certainty) of bodies of evidence. Our approach is to treat the whole mined patterns as a body of evidence. So, it is possible to use information measures to provide objective values that assist the experts in the evaluation and the interpretation of the mined patterns and, therefore, in the discovery of useful information from data.

In the best of our knowledge, the treatment of the mined patterns (characterized by its structure and frequency distribution) as a body of evidence is a novel approach that enable us to assess the information extracted by (temporal) data mining algorithms in a formal way. In this paper, we only used two information measures, but it is possible, and we propose this topic as a future work, to use additional information measures proposed in the literature to characterize mined bodies of evidence.

## REFERENCES

- Agrawal, R., Imielinski, T., and Swami, A. (1993). Mining association rules between sets of items in large databases. In *Proc. of the ACM SIGMOD Int. Conf. on Management of Data*, pages 207–216.
- Dubois, D. and Prade, H. (1985). A note on measures of specificity for fuzzy sets. *International Journal of General Systems*, 10:279–283.
- Dubois, D. and Prade, H. (1999). Properties of measures of information in evidence and possibility theories. *Fuzzy Sets and System*, 100:35–49.
- Guil, F., Bosch, A., and Marín, R. (2004). Tset: An algorithm for mining frequent temporal patterns. In *Proc. of the 1st Int. Workshop on Knowledge Discovery in Data Streams, in conjunction with ECML/PKDD 2004*, pages 65–74.
- Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ.
- Srikant, R. and Agrawal, R. (1996). Mining sequential patterns: Generalizations and performance improvements. In *Proc. of the 5th Int. Conf. on Extending Database Technology*, pages 3–17.
- Yager, R. (1981). Measurement of properties of fuzzy sets and possibility distributions. In *Proc. of the Third Int. Seminar on Fuzzy Sets*, pages 211–222.
- Yager, R. (1983). Entropy and specificity in a mathematical theory of evidence. *International Journal of General Systems*, 9:249–260.
- Zaki, M. (2001). Spade: An efficient algorithm for mining frequent sequences. *Machine Learning*, 42(1/2):31–60.