

What Makes Patterns Interesting in Knowledge Discovery Systems

Avi Silberschatz
Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ

Alexander Tuzhilin*
Information Systems Department
Stern School of Business
New York University

Abstract

One of the central problems in the field of knowledge discovery is the development of good measures of interestingness of discovered patterns. Such measures of interestingness are divided into *objective* measures – those that depend only on the structure of a pattern and the underlying data used in the discovery process, and the *subjective* measures – those that also depend on the class of users who examine the pattern. The focus of this paper is on studying subjective measures of interestingness. These measures are classified into *actionable* and *unexpected*, and the relationship between them is examined. The unexpected measure of interestingness is defined in terms of the *belief* system that the user has. Interestingness of a pattern is expressed in terms of how it affects the belief system. The paper also discusses how this unexpected measure of interestingness can be used in the discovery process.

Key Words: Measures of interestingness, patterns, actionability, unexpectedness, belief systems.

1 Introduction

It has been recognized early on in the knowledge discovery literature that a discovery system can generate a glut of patterns, most of which are of no interest to the user [5]. To address this problem, KDD researchers have been working on defining various *measures of interestingness* of patterns with the goal of developing KDD tools that discover only the patterns that are interesting according to these measures.

One approach to defining interestingness of a pattern is to define it in *objective* terms, where interestingness of a pattern is measured in terms of its structure and the underlying data used in the discovery process. For example, as Piatetsky-Shapiro points out in [11], the interestingness of a rule $A \rightarrow B$ is usually defined as a function of $p(A)$, $p(B)$ and $p(A \wedge B)$, where $p(\alpha)$ is the probability that condition α is true. A typical example of such objective measures of interestingness of a rule is the confidence and support measures [1].

*This work was supported in part by the NSF under Grant IRI-93-18773.

It has been noted in [12], however, that objective measures of interestingness, although useful in many respects, usually do not capture all the complexities of the pattern discovery process, and that *subjective* measures of interestingness are needed to define interestingness of a pattern. These subjective measures do not depend only on the structure of a rule and on the data used in the discovery process, but also on the *user* who examines the pattern. These measures recognize that a pattern that is of interest to one user, may be of no interest to another user. For example, a pattern discovering some security trading irregularities, such as insider trading, may be of great interest to the officials from the Securities and Exchange Commission (SEC). However, it is of very little use to a homeless person living in New York City.

In [12], subjective measures of interestingness were studied within the context of the discovery system KEFIR [10] that analyzes healthcare insurance claims for uncovering “key findings.” The *key findings* in KEFIR are statements about the most important deviations from the norms for various *indicators* assessing different characteristics of provision of healthcare, such as cost, usage, and quality. KEFIR defines interestingness of a key finding in terms of the estimated benefits (potential savings) of taking corrective action(s) that restore the deviation back to its norm. These corrective actions are specified in advance by the domain expert for various classes of deviations. This approach to defining a subjective measure of interestingness works nicely for KEFIR. However, it is very domain specific because 1) it deals only with patterns expressed as deviations, 2) it pre-classifies all the patterns that can be discovered into a finite (hopefully small) set of classes in order to assign a corrective action to each class (this is possible in case of KEFIR because it deals with a very domain-specific problem related to healthcare), and 3) it makes several domain-specific assumptions about the way estimated benefits are computed.

Klemettinen et al. [8] also observe that the objective measures of interestingness are not sufficient in many data mining applications because one can still generate a large number of strong rules that are interesting “objectively” but of little interest to the user. To address this problem, in [8] the authors propose a system of *templates*, that are rules expressed not in terms of the attributes of the data but in terms of the user-defined vocabulary [4] that is defined in terms of the data attributes. Then, a pattern (rule) is interesting, if it matches a restrictive template [8]. In their work, Klemettinen et al. bring the users into the discovery process by letting them specify the templates. However, [8] does not address the question of what the subjective measures of interestingness are and how they can be used in the discovery process.

In this paper, we focus on subjective measures of interestingness. However, unlike [12], we study them in a domain-independent context. In particular, we propose a classification of measures of interestingness and identify two major reasons why a pattern is interesting from the subjective

(user-oriented) point of view:

- *Unexpectedness* – a pattern is interesting if it is “surprising” to the user.
- *Actionability* – a pattern is interesting if the user can act on it to his advantage. This is, essentially, the subjective measure of interestingness studied in [12].

We also examine the relationship between these two measures of interestingness with the emphasis on the first measure.

2 Measures of Interestingness

As pointed out in the introduction, we identify two reasons why a pattern can be interesting to a user from the subjective point of view – *unexpectedness* and *actionability*. We now explain, at the intuitive level, what these two concepts mean and also explore how they are related to each other. We will base our discussions of these two subjective measures of interestingness on the following example which will be used throughout the paper.

Example 1 Consider a database of student evaluations of different courses offered at some university. Each semester, students do evaluations of all the courses they take, and the summaries of these evaluations are stored in the relation, the simplified form of which can be defined with the schema EVALUAT(TERM, YEAR, COURSE, SECTION, INSTRUCTOR, INSTRUCT_RATING, COURSE_RATING). □

2.1 Actionability Measure

According to this measure, a pattern is interesting because the user can *do* something about it; that is, the user can react to it to his or her advantage. For example, the pattern, *PAT*, that Professor X is consistently getting the overall instructor ratings below the overall course ratings ($\text{INSTRUCT_RATING} < \text{COURSE_RATING}$) can be of great interest to the chairperson of the department where X teaches because this shows to the chairperson that Professor X has “room” for improvement in his or her teaching ratings and X should be encouraged to work more on his teaching and presentation skills.

Actionability is an important subjective measure of interestingness because users are mostly interested in the knowledge that permits them to do their jobs better by taking some specific actions in response to the newly discovered knowledge. However, it is not the only important subjective measure of interestingness, as we discuss it in the next section.

2.2 Unexpectedness Measure

If a newly discovered pattern is surprising to the user, then it is certainly interesting (how many times we exclaimed “Oh, this is interesting...” when we discover something unexpected?). For example, if in most of the course evaluations, the overall instructor ratings (INSTRUCT_RATING) are *higher* than the overall course ratings (COURSE_RATING), and it turns out that in most of Professor X’s ratings overall instructor evaluations are *lower* than the overall course evaluations then such a pattern is unexpected and, hence, interesting. As another example, assume that in some course only 8% of the students responded with their evaluations, whereas this number is normally between 60% and 90%. This pattern (*RESP*) is definitely interesting because it is certainly unexpected.

We maintain that unexpected patterns are interesting because they contradict our expectations which, in turn, depend on our system of *beliefs*. For example, in the “instructor evaluation” pattern above, we believe that the overall instructor ratings should be higher than the overall course ratings, whereas the pattern contradicts this belief. Similarly, the “response rate” pattern contradicts our belief that there should be a “reasonable” response rate from the students.

2.3 Relationship Between Unexpectedness and Actionability

Clearly, some patterns are unexpected *and* actionable at the same time. For example, the “instructor evaluation” pattern *PAT* is actionable *and* unexpected at the same time.

Furthermore, some actionable patterns can be expected. For example, assume Professor Y is getting consistently high ratings, and the chairperson of the department wants to nominate him for the “Teacher of the Year” award. However before doing so, the chairperson wants to see the latest evaluations for Professor Y, and it turns out that they are also good. Such pattern (good student evaluations for Professor Y in the last semester) is expected. However, it is also actionable because the chair can nominate Professor Y for the award now without any reservations.

Also, a pattern can be unexpected and non-actionable. For example, the “response rate” pattern *RESP* is unexpected. However, it is not actionable for the chairperson of the department offering this course because student response rates are beyond his or her control and cannot be influenced by the chairperson in any way. Although the chairperson cannot do much about this pattern, nevertheless, it is of interest to him/her because the chairperson is curious to know what happened and why only 8% of the class responded.

Although the two types of interestingness are independent of each other, we believe that the

majority of actionable patterns are unexpected and that the majority of unexpected patterns are actionable. Therefore, we believe that unexpectedness is a good approximation for actionability and actionability is a good approximation for unexpectedness.

Furthermore, we believe that, although both actionability and unexpectedness are important, actionability is really *the key* concept that is of main interest to business people. However, actionability is an elusive concept that is very difficult to capture formally for the following reasons. First, we have to partition the space of all the patterns into a finite (and hopefully small) set of equivalence classes and associate an action or a class of actions with each equivalence class. Since in many cases we do not know the space of all the patterns, this task can simply be infeasible. Furthermore, even if we know this space, the partitioning it into equivalence classes and assignments of actions for each class can be a very complicated task. Secondly, even if we succeed in associating actions with classes of patterns, actions and the mapping of actions to patterns may often change over time, and the task of reassigning patterns to actions can be a very hard one. For these reasons, we believe that it is very difficult to capture actionability formally.

One way to address actionability is through unexpectedness since we believe that most actionable patterns are unexpected and vice versa. In Section 4, we propose a method to define interestingness of a pattern as a measure of its unexpectedness. Since we argued that most unexpected patterns are also actionable, this approach provides for a way to identify actionable patterns in addition to unexpected ones.

3 Unexpectedness and Beliefs

As indicated in Section 2.2, unexpectedness is related to beliefs, and therefore we have to study beliefs before defining unexpectedness. We will follow the approach to defining beliefs as logical statements (predicate formulae expressed in first-order logic) and assigning some *degree* or *measure*, or a *confidence factor* to each belief. For example, the degree of belief can be defined in Bayesian terms as the conditional probability that the belief holds given some previous “evidence” for this belief [6]. We describe this and some other approaches to defining degrees of beliefs later in this section. But first we classify beliefs into two types:

- **Hard Beliefs.** The hard beliefs are the *constraints* that cannot be changed with new evidence. In fact, if new evidence contradicts these beliefs, then there must be some mistakes made in acquiring this new evidence. For example, if it turns out that the number of students responding to the evaluation survey is greater than the number of students registered for that

class, then it means that the accuracy of this report is highly questionable. We would like to stress that hard beliefs are *subjective* and vary from one user to another. For example, for one user the statement that each person in the USA must have a unique Social Security Number is a hard belief, whereas, another user would be willing to question this belief.

- **Soft Beliefs.** These are beliefs that the user is willing to change with new evidence. For example, the belief that the overall instructor ratings are higher than the overall course ratings is a soft belief since it can be changed as new evidence (new grades) is reported every semester. Each soft belief is assigned a *degree* that specifies how strongly we believe in it.

Degrees of soft beliefs can be assigned in different ways, some of which are described below.

3.1 Bayesian approach

In this approach, the degree of a belief α is defined as a conditional probability, $P(\alpha|\xi)$, that α holds, given some previous evidence ξ supporting that belief. Given new evidence E , we update the degree of belief in α , $P(\alpha|E, \xi)$, using the Bayes rule:

$$P(\alpha|E, \xi) = \frac{P(E|\alpha, \xi)P(\alpha|\xi)}{P(E|\alpha, \xi)P(\alpha|\xi) + P(E|\neg\alpha, \xi)P(\neg\alpha|\xi)} \quad (1)$$

Let α_0 be the belief that “the overall instructor ratings are higher than the overall course ratings,” and assume that the degree of this belief is 0.85 ($P(\alpha_0|\xi) = 0.85$). Assume that a new course evaluation arrived containing a new evidence E_0 that the overall instructor rating (INSTRUCT_RATING) for the course taught by this instructor is 5.8 and the overall course rating (COURSE_RATING) is 5.2. We associate two functions with belief α_0 computing the conditional probabilities $P(E|\alpha_0, \xi)$ and $P(E|\neg\alpha_0, \xi)$ for evidence E . Assume that it turns out that for the new evidence E_0 , $P(E_0|\alpha_0, \xi) = 0.62$ and $P(E_0|\neg\alpha_0, \xi) = 0.42$. Then, substituting these numbers into (1), we compute $P(\alpha_0|E_0, \xi) = 0.89$.

3.2 Dempster-Shafer Approach

In Dempster-Shafer theory of evidential reasoning [14], one starts with a *frame of discernment* or Universe of Discourse (UoD) [16], denoted as Θ , that defines the domain of reference. Then a *basic probability assignment* is a mapping $m : 2^\Theta \rightarrow [0, 1]$ such that

$$\sum_{A \subseteq \Theta} m(A) = 1$$

The degree of belief given to $A \subset \Theta$ is specified by the *belief function*, $bel : 2^\Theta \rightarrow [0, 1]$, such that

$$bel(A) = \sum_{B \subset A, B \neq \emptyset} m(B)$$

In other words, the degree of a belief assigned to A is the sum of all the basic probability assignments $m(B)$ allocated to statements B that imply A .

Given two belief functions induced by two distinct pieces of evidence, the belief function that results from their combination is obtained by Dempster’s rule of combination [14].

3.3 Frequency approach

Assume that beliefs are defined as rules and that evidence ξ consists of either “raw” data (facts) or rules¹. To be specific, assume that “raw” data is stored in a relational database. Also, assume that belief $b \equiv \alpha \rightarrow \beta$ is safe [16] in the sense that all the variables appearing in β also appear positively in α among its free variables x_1, x_2, \dots, x_n .

Then we define the degree of belief b as follows. For evidence ξ consider the space \mathcal{D} of all the interpretations D that satisfy this evidence. For example, if ξ consists of “raw” data D , then there is only one interpretation (D) in this case. As another example, if ξ is the evidence that in Spring of 1996 men and women received equal grades in CS courses on average, then the space of interpretations consists of all the possible grades of the students who took CS courses in Spring of 1996, such that men’s and women’s grades in these courses were equal on average. Furthermore, we make an assumption that each of the interpretations $D \in \mathcal{D}$ is equi-probable.

To define the degree of belief, we compute for each interpretation $D \in \mathcal{D}$

$$d(b|D) = \frac{\text{Number of tuples } (x_1, x_2, \dots, x_n) \text{ satisfying } \alpha \text{ and } \beta \text{ on data } D}{\text{Number of tuples } (x_1, x_2, \dots, x_n) \text{ satisfying } \alpha \text{ on } D}$$

Then the degree of belief b is the expected value of $d(b|D)$ taken over all the interpretations $D \in \mathcal{D}$, i.e., $\mathcal{E}\{d(b|D)|D \in \mathcal{D}\}$.

3.4 Cyc’s Approach

Cyc [9] also considers beliefs and defines their degrees in terms of a *certainty factor (CF)*. A belief in Cyc is either a fact or a rule [9], and a certainty factor is either assigned to a belief by a knowledge editor, or is deduced by inferencing, or is computed from two pieces of evidence by combining certainty factors using simple heuristics described in [9].

¹In fact, facts can be viewed as degenerate rules of the form $\rightarrow \alpha$, where α is a ground atom.

Cyc provides two approaches to defining CF – numeric and logic-based. In numeric approach, the CF of a belief is a number between 0 and 100, whereas in the logic-based approach Cyc considers 5-valued logic and the “contention table” that combines certainty factors. In some sense, the logic-based approach can be perceived as a “discretized” numeric approach.

3.5 Statistical approach

In this approach, a belief is considered as a statistical hypothesis. For example, if we believe that men and women should receive equal grades in the courses they take, then such a belief is expressed as a null hypothesis $H_0 : \mu_m = \mu_w$, where μ_m and μ_w are average grades of men and women. Then the degree of a belief is defined as a *significance level* α [7] for which the statistic for the test is on the “border line” of acceptance of the hypothesis.

For example in case of the null hypothesis $H_0 : \mu_m = \mu_w$ assume that men’s and women’s grades are independent of each other. Then, if $\hat{\mu}_m$ and $\hat{\mu}_w$ are the observed sample means and $\hat{\sigma}_m$ and $\hat{\sigma}_w$ are the standard deviations for the men’s and women’s grades based on the actual data about the grades, and N_m and N_w are the number of men’s and women’s grades used in computing the means, then the degree of belief in H_0 is defined as follows. We first determine the point on the x -axis of the normal distribution $N(0, 1)$ corresponding to the value

$$z = \frac{|\hat{\mu}_m - \hat{\mu}_w|}{\sqrt{\frac{\hat{\sigma}_m^2}{N_m} + \frac{\hat{\sigma}_w^2}{N_w}}}$$

and then will find the significance level α that corresponds to that value of z [7].

One of the problems with the approach described in this section is that not any belief can be formulated as a testable hypothesis. For example, the hypothesis that men receive better grades than women cannot be tested in this way [7]. This means that we can define the degree of only *some* of the beliefs using the approach described in this section, which makes it impractical for the general case.

3.6 Comparison of the Approaches

The approaches to defining degrees of beliefs differ significantly from each other in terms of their generality, as well as the properties they satisfy. For example, the statistical approach cannot be applied to an arbitrary belief, as explained in Section 3.5. Similarly, the frequency approach presented in Section 3.3 is defined only for beliefs expressed as rules. In contrast to this, the Bayesian approach can be applied to arbitrary beliefs.

Similarly, the approaches differ from each other in terms of the properties they satisfy. For example, the Bayesian approach assumes that for any belief α , $d(\neg\alpha|\xi) + d(\alpha|\xi) = 1$, whereas the Dempster-Shafer approach does not assume this [14].

It has been argued by Cox [3] and Jaynes [6] (see also [2]) that the conditional (Bayesian) probability is the *only* appropriate approach to defining “degrees of plausibility” of logical statements satisfying certain intuitive desiderata (specified in [3] and [6]). For this reason, and because of its generality, it appears that the Bayesian approach is the most appropriate for defining degrees of beliefs among the approaches discussed in this section.

However, one of the problems with the Bayesian approach is that it is sometimes difficult to compute the posterior probability $P(\alpha|E, \xi)$ from its prior probability $P(\alpha|\xi)$ because of the problems with computing the conditional probabilities $P(E|\alpha, \xi)$ and $P(E|\neg\alpha, \xi)$ used in the Bayes formula. To determine these probabilities, one has to assign, for each belief α , the joint probability distribution $P_\alpha(E, \xi)$ over the space of all possible pairs of (ξ, E) , and this can be a formidable task.

To address this problem, these probabilities can be *estimated* using the following technique. We can select a set of parameters (or features) $\mathbf{x} = (x_1, x_2, \dots, x_n)$ for the application. For example, in the university application these parameters could be men’s and women’s average grades, x_m and x_w , in all the courses taught at that university over some period of time. Then we can estimate these parameters using the past evidence ξ . Let $\bar{\mathbf{x}} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n)$ be the vector of estimated means of these parameters based on the past evidence ξ and let $\mathbf{COV}(\mathbf{x})$ be the covariance matrix. Also, let E be the new evidence and $\hat{\mathbf{x}} = (x_1, x_2, \dots, x_n)$ be the estimators of the parameters $\mathbf{x} = (x_1, x_2, \dots, x_n)$ based on the new evidence E . For example, let E be the data containing grades received in all the courses during the current semester. Then we can make various assumptions about the probability distribution of $P(\hat{\mathbf{x}} | \bar{\mathbf{x}}, \mathbf{COV}(\bar{\mathbf{x}}))$ as a function of $\hat{\mathbf{x}}$. For example, we can assume that it is a multivariate normal distribution with the mean $\bar{\mathbf{x}}$ and the covariance matrix $\mathbf{COV}(\bar{\mathbf{x}})$. Then we can *approximate* the conditional probability $P_\alpha(E|\xi)$ with $P(\hat{\mathbf{x}} | \bar{\mathbf{x}}, \mathbf{COV}(\bar{\mathbf{x}}))$.

Although the Bayesian approach appears to be the most appropriate for defining the degree of belief among the approaches considered in this section, the definition of the measure of interestingness defined in the next section does not depend on how the degree of belief is specified. Therefore, any degree of belief introduced in this section can be used in that definition.

4 Interestingness of Patterns

As we stated in Section 2.2, intuitively, a pattern is interesting relative to some belief system if it “affects” this system, and the more it “affects” it, the more interesting the pattern is. We distinguish between hard and soft beliefs, and thus treat these two cases separately.

Hard Beliefs. If a pattern contradicts the set of hard beliefs of the user then this pattern is *always* interesting to the user. Note that a contradicting pattern does not affect hard beliefs and that hard beliefs are kept unchanged. Such a contradiction means that the data used to derive the pattern must be wrong. For example, in the “response rate” pattern described in Section 3, something must have been wrong with the data because the number of responding students cannot be greater than the number of students registered for the course.

Soft Beliefs. We pointed out already that, intuitively, the more a pattern “affects” the belief system, the more interesting it is. Since different beliefs vary in their importance, changes to some beliefs should be more interesting than changes to others. To capture this formally, we associate a weight function w_i with each soft belief α_i in the belief system B . In addition, we normalize weights to 1, that is, $\sum_{\alpha_i \in B} w_i = 1$. We formally define interestingness of pattern p relative to a (soft) belief system B and previous evidence ξ as

$$I(p, B, \xi) = \sum_{\alpha_i \in B} w_i |d(\alpha_i|p, \xi) - d(\alpha_i|\xi)| \quad (2)$$

This definition of interestingness measures by how much degrees of beliefs changed as a result of a new pattern p . As an important special case, we can assume that all the beliefs in the belief system carry equal weights (i.e., $w_i = 1/N$, where N is the number of beliefs in B).

Example 2 Assume that the degree of a belief is measured with the conditional probability as described in Section 3.1 and assume that B consists of only one belief α_0 that “the overall instructor ratings are higher than the overall course ratings.” In that section it was shown how $P(\alpha_0|\xi) = 0.85$ was changed to $P(\alpha_0|E_0, \xi) = 0.89$ with new evidence E_0 that in the new faculty-course evaluation report for Professor X the instructor rating (5.8) was higher than the course rating (5.2). Substituting these numbers into (2), we obtain the interestingness of the pattern E_0 \square

If the degree of a belief is defined in Bayesian terms, then it satisfies the following important properties. First, it follows from the fact that if we replace a belief α in a belief system B with its negation $\neg\alpha$ and denote the resulting belief system B' , then $I(p, B, \xi) = I(p, B', \xi)$. This observation is summarized in the following proposition.

Proposition 1 *An interestingness of a pattern relative to a belief system B does not change if any belief(s) in B are replaced by its (their) complements.*

The next proposition establishes an important link between the measure of interestingness, as defined by (2), and the concept of unexpectedness, described in Section 2.2.

Proposition 2 *Let α be a belief, such that $0.5 < d(\alpha|\xi) < 1$. Let p be a pattern confirming belief α ($\alpha \models p$) such that $0.5 < d(\alpha|\neg p, \xi)$ and $d(\alpha|p, \xi) < 1$. Then $I(p, \alpha, \xi) \leq I(\neg p, \alpha, \xi)$ ².*

In other words, Proposition 2 shows that unexpected patterns are more interesting than expected patterns.

5 Interestingness and the Discovery Process

Once we define interestingness of a pattern, we can search for patterns that are interesting according to this measure. In this section, we discuss one search method that relies on the concept of interestingness introduced in Section 4.

The most interesting and challenging discovery problem arises when the data changes over time because the patterns also keep changing with the data. This is a typical situation in On-Line Transaction Processing (OLTP) systems, such as airline reservations, banking, and insurance claim processing systems. Therefore, we will focus on an environment where new data is periodically added to the previously collected historical data.

When a new set of data is added to a database, it affects the degrees of beliefs about the data stored in that database. One way of searching for interesting patterns in the data is to examine the changes in degrees of beliefs resulting from adding new data. The following proposition shows that, if degrees of some of the beliefs change, this means that the newly arrived data contains some interesting patterns.

Let D be the old (historical) data stored in a database, and ΔD be the new data that was just added to D . Let B be a belief system about the data stored in the database.

Proposition 3 *If there is a belief α in B such that $d(\alpha|\Delta D, D) \neq d(\alpha|D)$, then there exists a pattern p in ΔD such that $I(p, B, D) \neq 0$.*

²Note that pattern p is expected in this context because it follows from α and $\neg p$ is unexpected because it contradicts α .

This proposition suggests the following *belief-driven discovery* scheme. When new data arrives, degrees of all the beliefs are revised based on the new data. If some of the degrees change above predetermined threshold levels, this means that there are some interesting patterns in the data and that the discovery processes to extract interesting patterns should be launched. Description of such processes is beyond the scope of this paper and they are described in [15]. In this paper, we only provide a theoretical justification for this discovery scheme based on Proposition 3.

6 Conclusions

Measures of interestingness of patterns in data mining applications can be classified into objective and subjective. We classified subjective measures into unexpected and actionable and argued, at the intuitive level, that these two measures of interestingness are independent of each other. Although actionability appears to be the major concept, we believe that it is a difficult notion to capture formally. Since we believe that most unexpected patterns are actionable and most actionable patterns are unexpected, we proposed in the paper to capture actionability via unexpectedness. Therefore, we studied “unexpectedness” as a measure of interestingness and defined interestingness of a pattern in terms of how strongly it “shakes” the existing system of beliefs. We also demonstrated that this definition makes unexpected patterns more interesting than the expected ones.

Once we defined a measure of interestingness based on unexpectedness, we can use it to discover interesting patterns in the data. We are currently working on one discovery scheme that uses the system of beliefs and changes to their degrees to discover interesting patterns. The initial report describing this *belief-driven discovery* method is presented in [15]. We are also working on the problem of building and maintaining belief systems using the *data-monitoring and discovery-triggering* paradigm presented in [13].

References

- [1] R. Agrawal, T. Imielinsky, and A. Swami. Mining association rules between sets of items in large databases. In *Proceedings of ACM SIGMOD Conference*, pages 207–216, 1993.
- [2] P. Cheeseman. In defense of probability. In *Proceedings of the IJCAI Conference*, 1985.
- [3] R. T. Cox. On inference and inquiry – an essay in inductive logic. In Levine and Tribus, editors, *The Maximum Entropy Formalisms*. M.I.T. Press, 1979.

- [4] V. Dhar and A. Tuzhilin. Abstract-driven pattern discovery in databases. *IEEE Transactions on Knowledge and Data Engineering*, 5(6), 1993.
- [5] W.J. Frawley, G. Piatetsky-Shapiro, and C.J. Matheus. Knowledge discovery in databases: an overview. In G. Piatetsky-Shapiro and W.J. Frawley, editors, *Knowledge Discovery in Databases*. AAAI / MIT Press, 1991.
- [6] E.T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press. To appear.
- [7] S.K. Kachigan. *Statistical Analysis*. Radius Press, 1986.
- [8] M. Klemettinen, H. Mannila, P. Ronkainen, H. Toivonen, and A. I. Verkamo. Finding interesting rules from large sets of discovered association rules. In *Proceedings of the Third International Conference on Information and Knowledge Management*, December 1994.
- [9] D. Lenat and R.V. Guha. *Building Large Knowledge-Based Systems*. Addison-Wesley, 1990.
- [10] C. J. Matheus, G. Piatetsky-Shapiro, and D. McNeill. An application of KEFIR to the analysis of healthcare information. In *Proceedings of the AAAI-94 Workshop on Knowledge Discovery in Databases*, 1994.
- [11] G. Piatetsky-Shapiro. Discovery, analysis, and presentation of strong rules. In G. Piatetsky-Shapiro and W.J. Frawley, editors, *Knowledge Discovery in Databases*. AAAI / MIT Press, 1991.
- [12] G. Piatetsky-Shapiro and C. J. Matheus. The interestingness of deviations. In *Proceedings of the AAAI-94 Workshop on Knowledge Discovery in Databases*, pages 25–36, 1994.
- [13] A. Silberschatz and A. Tuzhilin. User-assisted knowledge discovery: How much should the user be involved. In *Proceedings of the SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery*, Montreal, June 1996.
- [14] P. Smets. Belief functions. In P. Smets, A. Mamdani, D. Dubois, and Prade H., editors, *Non-Standard Logics for Automated Reasoning*. Academic Press, 1988.
- [15] A. Tuzhilin and A. Silberschatz. A belief-driven discovery framework based on data monitoring and triggering. In preparation.
- [16] J. Ullman. *Principles of Database and Knowledge-Base Systems*, volume 1. Computer Science Press, 1988.