A Step Towards the Foundations of Data Mining

Y.Y. Yao
Department of Computer Science, University of Regina
Regina, Saskatchewan, Canada S4S 0A2

ABSTRACT

This paper addresses some fundamental issues related to the foundations of data mining. It is argued that there is an urgent need for formal and mathematical modeling of data mining. A formal framework provides a solid basis for a systematic study of many fundamental issues, such as representations and interpretations of primitive notions of data mining, data mining algorithms, explanations and applications of data mining results. A multi-level framework is proposed for modeling data mining based on results from many related fields. Formal concepts are adopted as the primitive notion. A concept is jointly defined as a pair consisting of the intension and the extension of the concept, namely, a formula in a certain language and a subset of the universe. An object satisfies the formula of a concept if the object has the properties as specified by the formula, and the object belongs to the extension of the concept. Rules are used to describe relationships between concepts. A rule is expressed in terms of the intensions of the two concepts and is interpreted in terms of the extensions of the concepts. Several different types of rules are investigated. The usefulness and meaningfulness of discovered knowledge are examined using a utility model and an explanation model.

Keywords: Foundations of data mining, rule mining, formal concepts, explanation oriented mining, data mining models

1. INTRODUCTION

Data mining may be simply defined as the extraction or mining of knowledge from large amounts of data.[10] This simple definition can be further elaborated by clearly specifying features of data mining tasks and the discovered knowledge. For example, according to Frawley *et al.*,[6] data mining can be viewed as "the non-trivial extraction of implicit, previously unknown and potentially useful information from data". Berry and Linoff[1] defined data mining as "the process of exploration and analysis, ..., of large quantities of data in order to discover meaningful patterns and rules". From the two definitions, one can observe many fundamental issues of data mining, such as data and knowledge (information), data mining tasks, data mining algorithms, and data mining results. Although the two definitions differ slightly with respect to those issues, they both rely on our intuitive and common-sense understanding of notions such as data, knowledge, usefulness and meaningfulness.

Extensive studies of data mining can be roughly classified into a few streams, namely, foundations of data mining, data mining process, data mining algorithms and methodologies (include all algorithms in every phrase of the data mining process), evaluation, interpretation and visualization of data mining results, and applications of data mining results.[5, 10] Unfortunately, these streams have not been developed evenly. Extensive studies in the field have been focused on algorithms and methodologies for mining different types of knowledge, speeding up of existing algorithms, and evaluation of discovered knowledge.[5] There is very little attention being paid to the foundations of data mining, and in particularly, to the formal and mathematical modeling of data mining.[16, 29] With the maturity of data mining algorithms and techniques, the foundations of data mining have received much attention lately.[3, 8, 9, 13, 14, 28, 29]

This paper is an attempt towards the foundations of data mining. It extends our initial investigations to a wider context.[29] In Section 2, we attempt to clarify the terms "foundations" and "foundations of data mining" by following Simpson's essay on foundations of mathematics.[21] In Section 3, a multi-level framework is proposed for modeling data mining. The lower (inner) levels concentrate on philosophical and conceptual development, the middle levels on data mining algorithms, and the upper (outer) levels on interpretations and applications of data mining results.

E-mail: yyao@cs.uregina.ca, http://www.cs.uregina.ca/∼yyao

2. FOUNDATIONS OF DATA MINING

As pointed out by Mannila,[16] the study of foundations of data mining is in its infancy, and there are probably more questions than answers. Although many data mining related conferences list foundations of data mining as one of the topics, there still does not exist a well accepted definition. The first step is therefore to make clear the meaning of foundations of data mining.

2.1. Review of existing approaches

Different approaches have been used in an attempt to establish the foundations of data mining. Some authors focus on philosophical studies of data mining. They ask fundamental questions such as "What is data mining? What is the scope of data mining? What is the difference between data mining and statistics?"[8, 11] Some authors view the foundations of data ming as the theory of data mining, and hence use the term "theoretical foundations of data mining".[10, 16] Mannila[16] put forward a convincing argument for the needs of theory of data mining by looking at the role played by the relational model in the development of databases. Theories that can be used to form a basis of data mining include data reduction, data compression, pattern discovery, probability theory, microeconomic view, and inductive databases.[10, 16] Some author advocate the needs for a unified framework and mathematical modeling of data mining.[9, 12, 29] A formal model would provide a common ground on which various methods can be studied and compared.[29] Hopefully, a well accepted model can provide a common interpretation for many basic notions that have been either defined or named differently by researchers.

Foundations of data mining can be examined either on abstract and philosophical levels or on concrete and technical levels. An example of each is reviewed below. On the philosophical level, Chen[3] suggested that foundations of data mining can be studied from three different but related dimensions, the philosophical dimension, the technical dimension, and the social dimension. The philosophical dimension deals with the nature and scope of data mining. The technical dimension covers data mining methods and techniques. The social dimension concern the social impact and consequences of data mining. In general, one may consider a multi-dimensional view if new dimensions are added and existing dimensions are further divided. On the technical level, Xie and Raghavan[28] developed a logical foundation of data mining based on Bacchus' probability logic. Their work points at an important direction in building foundations of data mining, namely, the precise definition of intuitive notions. For example, they precisely defined "pattern", "previously unknown knowledge" and "potentially useful knowledge" in the model. Furthermore, a logic induction operator is defined, which can be used to discover "previously unknown and potentially useful knowledge".

2.2. Foundations and Foundations of Data Mining

While it is easy to argue for the needs of studies on foundations of data mining, it is not so easy to state the goals, scopes, and methodologies of such studies. Nevertheless, a promising starting point is a clarification of the meanings of the terms "foundations" and "foundations of data ming".

Simpson[21] gave a concise and in-depth discussion of the term "foundations" for the study of the foundations of mathematics. The relevant paragraphs are quoted here:

> All human knowledge is conceptual and forms an integrated whole. All human knowledge is contextual (the context is the entire sum of human knowledge, which must be consistent) and hierarchical (organized in a tower or actually a partial ordering, where the base or minimal elements of the ordering are the most fundamental concepts, and higher-level concepts depend on lower-level concepts).
>
> Within the integrated whole of human knowledge, we may focus on what are usually called "fields of study" or "specialties". A field of study is distinguished by a certain conceptual unity: the concepts that make up the field are closely related to each other and are sufficiently self-contained so that the field lends itself to study in isolation for some purposes. Usually if not always, this kind of conceptual unity follows from the existence of a specific subject matter, the real-world object of study.
>
> If X is any field of study, "foundations of X" refers to a more-or-less systematic analysis of the most basic or fundamental concepts of field X. The term "basic" or "fundamental" here refers to the natural ordering or hierarchy of concepts (see point 1 [first quoted paragraph] above).

> In the history of particular fields of study, the foundations often take time to develop. At first the concepts and their relationships may not be very clear, and the foundations are not very systematic. As time goes on, certain concepts may emerge as more fundamental, and certain principles may become apparent, so that a more systematic approach becomes appropriate.
>
> The foundations of X are not necessarily the most interesting part of field X. But foundations help us to focus on the conceptual unity of the field, and provide the links which are essential for applications and for integration into the context of the rest of human knowledge.
>
> What is properly regarded as foundational in one context should not necessarily be regarded as foundational in another.

In accordance with Simpson's description, the foundations of a field can be built by studying the fundamental elements of the fields and the inherent hierarchial structure of the field. The fundamental elements are primitive notions, principles, and axioms, on which other concepts and results are based. A clear and well accepted hierarchical organization of a field can be expected with the maturity of the field. Furthermore, each branch or subfield of a field may have its own foundational elements and hence its own foundations.

Data mining may be viewed as a field on its own. Foundations of data mining means a systematic study of various notions that form its inherent hierarchical structure, from the basic concepts (data, objects, attribute/features, knowledge, etc.) to the theories, methodologies and algorithms for deriving knowledge from data (data mining algorithms, data mining process, etc.), and to the evaluation and interpretation of data mining results (visualization, interestingness, profit,utility, actionability, etc.). One may propose a formal and unified framework for this purpose. The properties and limitations of the formal framework need to be carefully examined.

Data mining may also be viewed as an interdisciplinary study involving many fields such as statistics, cognitive science, machine learning, and artificial intelligence. Foundations of data mining may be approached by an examination of the foundations of the related fields. Similar to the study on foundations of cognitive science by Simon and Kaplan,[20] one needs to examine the goal of data mining, the principal contributing disciplines to data mining, the architecture of data mining systems, and methods of data mining researches. It is necessary to identify a range of topics that help define the goal and scope of data mining and numerous dimensions along which the field can be explored. The results from studies on foundations of knowledge acquisition more than ten years ago can be immediately applied to the study of foundations of data mining.[2]

3. A MULTI-LEVEL FRAMEWORK FOR MODELING DATA MINING

A data mining system may be viewed as an intermediate system between a database and an application, whose main task is to change data into useable knowledge. The theory of databases provides a solid basis for storing, manipulating, and retrieving data, at syntactical, symbolic levels. Knowledge is an entity in the semantic levels of data. Knowledge embedded in data is related to semantic interpretations of data.

The existence of knowledge in data is unrelated to whether we have an algorithm to extract it.[3] This observation has a significant implication for the study of foundations of data mining. Many studies on data mining reply on the existence of an algorithm, rather than the existence of knowledge. Even worse, more often than not, one implicitly assumes that a particular type of knowledge is useful simply because we have found an algorithm. It is therefore not surprising that we have many algorithms mining "easily extractable knowledge", which may not necessarily be interesting and useful.

In building a framework for data mining, it becomes obvious that we need to separate the study of knowledge and the study of data mining algorithms, and in turn to separate them from the study of utility of discovered knowledge. To achieve this goal, we propose a multi-level framework. The kernel focuses on the study of knowledge without reference to data mining algorithms. The technique levels focus on data mining algorithms without reference to particular application. The application levels focus on the utility of discovered knowledge with respect to particular domains of applications.

3.1. Kernel

In the kernel (inner levels), we study the philosophical issues and primitive notions of data mining. It is necessary to define precisely "knowledge" and the "basic unit of knowledge", which serve as the primitive notions of data mining. The definition is independent of any algorithms for mining the knowledge. Similarly, it is necessary to describe and classify different types of knowledge.

3.1.1. concepts as a primitive notion of data mining

We choose "concepts" or "formal concepts" as a primitive notion of data mining, which precisely defines "knowledge".[29] The justification for this choice can be drawn from many established fields. The study of concepts is central to philosophy, psychology, cognitive science, inductive data processing and analysis, and inductive learning.[17, 22, 23, 25] For example, Simth[22] states,

> In cognitive psychology and philosophy of mind, concepts are assumed to be the basic constituents of thought and belief. In linguistics the study of word meanings or lexical representations often involves the study of those concepts that are coded by single words. And in artificial intelligence, more often than not, proposals about knowledge representation *are* proposals about concepts. Moreover in each of these disciplines concepts are intimately linked to the process of drawing inductive inferences.

Inductive concept learning also plays a major role in machine learning.[17] The relevance of concepts to inductive data analysis has been demonstrated by many studies.[17, 22, 23, 25] Thus, if one views data mining as an inductive inference or learning process,[16, 28] the choice of concepts as a primitive notion is particularly meaningful.

Another reason for choosing concepts is that it is a very rich notion with many interpretations.[22, 25] The classical view treats concepts as entities with well-defined borderlines and describable by sets of singly necessary and jointly sufficient conditions.[25] Other views include the prototype view, the exemplar view, the frame view, and the theory view.[25] The applications of different views for inductive data analysis have also been addressed by many authors.[19, 23, 25]

For simplicity, we adopt the classical view of concepts. Every concept is understood as a unit of thoughts that consists of two parts, the intension and the extension of the concept.[4, 22, 23, 25, 27] The intension (comprehension) of a concept consists of all properties or attributes that are valid for all those objects to which the concept applies. The extension of a concept is the set of objects or entities which are instances of the concept. All objects in the extension have the same properties that characterize the concept. In other words, the intension of a concept is an abstract description of common features or properties shared by elements in the extension, and the extension consists of concrete examples of the concept. A concept is thus described jointly by its intension and extension. This formulation enables us to study concepts in a logic setting in terms of intensions and also in a set-theoretic setting in terms of extensions.

In order to obtain a computational model and to be consistent with data mining practice, we adopt Tarski's approach to study concepts through the notions of a model and satisfiability.[4, 18, 29]

Let U be a set of universe whose elements are called objects, and At a set of attributes or features. For each attribute $a \in At$, we associate it with a set of values or labels V_a. Furthermore, for each $a \in At$, there is mapping I_a connecting elements of U and elements of V_a. It should be pointed out that elements of U may not necessarily correspond to some physical objects. The same can be said about At and V_a. With respect to a dataset, we can build a model:

$$M = (U, At, \{V_a \mid a \in At\}, \{I_a \mid a \in At\}). \tag{1}$$

In the model, we consider a language $\mathcal{L}$ used to formally define concepts. In the language, we have a set of atomic formulas corresponding to some basic concepts. For each atomic formula α, it is assumed that an object x either satisfies α, written $x \models \alpha$, or does not satisfy α, written $\neg x \models \alpha$. From atomic formulas, we can construct other formulas by applying the logic connectives $\neg$, $\wedge$, $\vee$, $\rightarrow$, and $\leftrightarrow$. The satisfiability of any formula is defined as follows:

$$
\begin{array}{ll}
(1) & x \models \neg\phi \text{ iff } \text{ not } x \models \phi, \\
(2) & x \models \phi \wedge \psi \text{ iff } x \models \phi \text{ and } x \models \psi, \\
(3) & x \models \phi \vee \psi \text{ iff } x \models \phi \text{ or } x \models \psi, \\
(4) & x \models \phi \rightarrow \psi \text{ iff } x \models \neg\phi \vee \psi, \\
(5) & x \models \phi \leftrightarrow \psi \text{ iff } x \models \phi \rightarrow \psi \text{ and } x \models \psi \rightarrow \phi.
\end{array}
$$

If ϕ is a formula, the set $m(\phi)$ defined by:

$$m(\phi) = \{x \in U \mid x \models \phi\}, \tag{2}$$

is called the meaning of the formula ϕ in M. The meaning of a formula ϕ is therefore the set of all objects having the property expressed by the formula ϕ. In other words, ϕ can be viewed as the description of the set of objects $m(\phi)$. Thus, a connection between formulas and subsets of U is established. Obviously, the following properties hold[18]:

$$
\begin{array}{ll}
\text{(a)} & m(\alpha) = \{x \in U \mid x \text{ satisfies } \alpha\}, \\
\text{(b)} & m(\neg\phi) = -m(\phi), \\
\text{(c)} & m(\phi \wedge \psi) = m(\phi) \cap m(\psi), \\
\text{(d)} & m(\phi \vee \psi) = m(\phi) \cup m(\psi), \\
\text{(e)} & m(\phi \rightarrow \psi) = -m(\phi) \cup m(\psi), \\
\text{(f)} & m(\phi \equiv \psi) = (m(\phi) \cap m(\psi)) \cup (-m(\phi) \cap -m(\psi)).
\end{array}
$$

With the introduction of language $\mathcal{L}$, we have a formal description of concepts. A concept definable in a model M is a pair $(\phi, m(\phi))$, where $\phi \in \mathcal{L}$. More specifically, ϕ is a description of $m(\phi)$ in M, the intension of concept $(\phi, m(\phi))$, and $m(\phi)$ is the set of objects satisfying ϕ, the extension of concept $(\phi, m(\phi))$.

A concept $(\phi, m(\phi))$ is said to be a sub-concept of another concept $(\psi, m(\psi))$, or $(\psi, m(\psi))$ a super-concept of $(\phi, m(\phi))$, if $m(\phi) \subseteq m(\psi)$. A concept $(\phi, m(\phi))$ is said to be a smallest non-empty concept in M if there does not exist another non-empty proper sub-concept of $(\phi, m(\phi))$. Two concepts $(\phi, m(\phi))$ and $(\psi, m(\psi))$ are disjoint if $m(\phi) \cap m(\psi) = \emptyset$. If $m(\phi) \cap m(\psi) \neq \emptyset$, we say that the two concepts have a non-empty overlap and hence are related.

We can also form a concept lattice based on logical implication $\rightarrow$ or set inclusion $\subseteq$. More specifically, for two concepts $(\phi, m(\phi))$ and $(\psi, m(\psi))$, the meet $\sqcap$ and join $\sqcup$ are defined by:

$$
\begin{aligned}
(\phi, m(\phi)) \sqcap (\psi, m(\psi)) &= (\phi \wedge \psi, m(\phi) \cap m(\psi)), \\
(\phi, m(\phi)) \sqcup (\psi, m(\psi)) &= (\phi \vee \psi, m(\phi) \cup m(\psi)).
\end{aligned} \tag{3}
$$

One can easily define the extension based on the intension of a concept. The reverse is not necessarily true. For the same extension, one may find totally two different intensions in the model M. This implies that knowledge inferred from a particular dataset may not always be meaningful.

Based on the notions introduced so far, we can study a special type of knowledge represented by relationship between overlap concepts. This type of knowledge is commonly referred to as rules. A rule can be expressed in the form, $\phi \Rightarrow \psi$, where ϕ and ψ are intensions of two concepts. A crucial issue is therefore the characterization, classification, and interpretation of rules. It is reasonable to expect that different types of rules represent different kinds of knowledge derivable from a database. These issues are addressed in the following two subsections.

3.1.2. Rules and their probabilistic characterizations

In data mining, rules are typically interpreted in terms of conditional probability.[31] For a rule $\phi \Rightarrow \psi$, its characteristics can be summarized by the following contingency table:

	ψ	$\neg\psi$	Totals
ϕ	a	b	$a+b$
$\neg\phi$	c	d	$c+d$
Totals	$a+c$	$b+d$	$a+b+c+d=n$

$$a = |m(\phi \wedge \psi)|, \qquad b = |m(\phi \wedge \neg\psi)|,$$
$$c = |m(\neg\phi \wedge \psi)|, \qquad d = |m(\neg\phi \wedge \neg\psi)|.$$

Different measures can be defined to reflect various aspects of rules.

The *generality* of ϕ is defined by:

$$G(\phi) = \frac{|m(\phi)|}{|U|} = \frac{a+b}{n}, \tag{4}$$

which indicates the relative size of the concept ϕ. Obviously, we have $0 \leq G(\phi) \leq 1$. A concept is more general if it covers more instances of the universe. A sub-concept has a lower generality than its super-concept. The quantity may be viewed as the probability of a randomly selected element satisfying ϕ.

The *absolute support* of ψ provided by ϕ is the quantity:

$$AS(\phi \Rightarrow \psi) = AS(\psi|\phi) = \frac{|m(\psi) \cap m(\phi)|}{|m(\phi)|} = \frac{a}{a+b}, \tag{5}$$

The quantity, $0 \leq AS(\psi|\phi) \leq 1$, states the degree to which ϕ supports ψ. It may be viewed as the conditional probability of a randomly selected element satisfying ψ given that the element satisfies ϕ. In set-theoretic terms, it is the degree to which $m(\phi)$ is included in $m(\psi)$. Clearly, $AS(\psi|\phi) = 1$, if and only if $m(\phi) \neq \emptyset$ and $m(\phi) \subseteq m(\psi)$. That is, a rule with the maximum absolute support 1 is a certain rule.

The *mutual support* of ϕ and ψ is defined by:

$$MS(\phi, \psi) = \frac{|m(\phi) \cap m(\psi)|}{|m(\phi) \cup m(\psi)|} = \frac{a}{a+b+c}. \tag{6}$$

One may interpret the mutual support, $0 \leq MS(\phi, \psi) \leq 1$, as a measure of the strength of a pair of rules $\phi \Rightarrow \psi$ and $\psi \Rightarrow \phi$.

The *change of support* of ψ provided by ϕ is defined by:

$$CS(\phi \Rightarrow \psi) = CS(\psi|\phi) = AS(\psi|\phi) - G(\psi) = \frac{a}{a+b} - \frac{a+c}{n}. \tag{7}$$

Unlike the absolute support, the change of support varies from -1 to 1. One may consider $G(\psi)$ to be the prior probability of ψ and $AS(\psi|\phi)$ the posterior probability of ψ after knowing ϕ. The difference of posterior and prior probabilities represents the change of our confidence regarding whether ϕ actually related to ψ. For a positive value, one may say that ϕ positively related to ψ; for a negative value, one may say that ϕ is negatively related to ψ.

The generality $G(\psi)$ is related to the satisfiability of ψ by all objects in the database, and $AS(\phi \Rightarrow \psi)$ is related to the satisfiability of ψ in the subset $m(\phi)$. A high $AS(\phi \Rightarrow \psi)$ does not necessarily suggest a strong association between ϕ and ψ, as a concept ψ with a large $G(\psi)$ value tends to have a large $AS(\phi \Rightarrow \psi)$ value. The change of support $CS(\phi \Rightarrow \psi)$ may be more accurate.

3.1.3. Association rules, exception rules and peculiarity rules

Association rules were first proposed for mining rules from transaction databases. A transaction database can be equivalently represented by the following model:

$$M = (U, At, \{V_a \mid a \in At\}, \{I_a \mid a \in At\}), \quad \text{where}$$

$$
\begin{aligned}
&U = T = \text{a set of transactions;}\\
&At = I = \text{a set of items;}\\
&V_i = \{0, 1\}, i \in I;\\
&I_i(t) = \begin{cases} 0, & \text{transaction } t \text{ does not contain item } i,\\ 1, & \text{transaction } t \text{ contains item } i. \end{cases}
\end{aligned}
$$

The language $\mathcal{L}$ is defined as follow. For each item $i \in I$, we have an atomic formula α_i. A transaction t satisfies α_i if it contains i, namely, $I_i(t) = 1$, and otherwise it does not satisfy i, namely, $I_i(t) = 0$. The meaning of α_i is therefore given by:

$$m(\alpha_i) = \{t \in T \mid I_i(t) = 1\}. \tag{8}$$

In association mining, one only considers formulas constructed by the logic connective $\wedge$. One can easily extend association rules to non-transaction databases and allows more general formulas in the rules.[10]

Two measures are used for mining association rules of the form $\phi \Rightarrow \psi$. They are the generality of $\phi \wedge \psi$ and the absolute support $AS(\phi \Rightarrow \psi)$. By specifying threshold values of generality and absolute support, one can obtain all association rules whose generality and absolute support are above the thresholds. With an association rule, it is very tempting to relate a large absolute support with a strong association between two concepts. However, such a connection may not exist. Suppose we have $AS(\phi \Rightarrow \psi) = 0.90$. If we also have $G(\psi) = 0.95$, we can conclude that ϕ is in fact negatively associated with ψ. This suggests that an association rule may not reflect the true association. Conversely, an association rule with low generality may have a large change of support. In mining association rules, concepts with low generality are not considered in the search for association. On the other hand, two concepts with low generality may have either large absolute support or a large change of support. Unfortunately, traditional association mining fail to find such useful rules.

Exception rules have been studied as extension of association rules to resolve some of the above problems.[24] For an association rule, $\phi \Rightarrow \psi$, with high absolute support, one may associate an exception rule $\phi \wedge \phi' \Rightarrow \neg\psi$. Roughly speaking, ϕ' can be viewed as the condition for exception to rule $\phi \Rightarrow \psi$. To be consistent with the intended interpretation of exception rule, it is reasonable to assume that $\phi \wedge \phi' \Rightarrow \neg\psi$ have a high absolute support and a low generality. More specifically, we would expect a low generality of $\phi \wedge \phi'$. Otherwise, the rule cannot be viewed as describing exceptional situations.

Zhong *et al.*[32] identified and studied a new class of rules called peculiarity rules. In mining peculiarity rules, the distribution of attribute values is taken into consideration. Attention is paid to objects whose attribute values are different from that of other objects. After the isolation of such peculiarity data, rules with low generality and high absolute support, and consequently a high change of support, are searched. Although a peculiarity rule may share the same properties with an exception rule, as expressed in terms of generality, absolute support, and change of support, it does not express exception to another rule.

We can qualitatively characterize association rules, exception rules and peculiarity rules in the following table:

Rule	G (Generality)	AS (Absolute support)	CS (Change of Support)
Association rule:			
$\phi \Rightarrow \psi$	High	High	Unknown
Exception rule:			
$\phi \Rightarrow \psi$	High	High	Unknown
$\phi \wedge \phi' \Rightarrow \neg\psi$	Low	High	High
Peculiarity rule:			
$\phi \Rightarrow \psi$	Low	High	High

Both exception rule mining and peculiarity rule mining aim at finding rules missed by association rule mining. While exception rules and peculiarity rules have a high change of support, indicating a strong association between two concepts, association rules do not necessarily have this property. All three types of rules focus on high level of absolute support. For an exception rule, it is also expected that the generality of $\phi \wedge \phi'$ is low. For a peculiarity rule, the generalities of both ϕ and ψ are expected to be low. In contrast, the generality of right hand of an exception rule does not have to be low.

One may say that rules with high absolute support and high change of support are of interest. The use of generality in association rule mining is mainly for the sake of low computational cost, rather than semantics consideration. Exception rules and peculiarity rules are two subsets of rules with high absolute support and high change of support. It may be interesting to design algorithms to find *all* rules with high absolute support and/or high change of support. All three types of rules are not related to the mutual support measure MS. In some application, it may be necessary to consider rules $\phi \Rightarrow \psi$ and $\psi \Rightarrow \phi$ together.

3.1.4. Additional types of knowledge in data

The potential usefulness of a formal framework has been demonstrated in the previous subsections with respect to the representation and analysis of different types of rules. The results of the analysis clearly show the similarities and differences between many existing data mining methods, as well as the limitation of each one. We can further demonstrate the usefulness of the model by identifying more types of knowledge embedded in data.

- **Similarity**: Two concepts ϕ and ψ are considered to be similar if the mutual support $MS(\phi, \psi)$ is high.
- **Negative association**: Consider two concepts ϕ and ψ such that both $G(\phi)$ and $G(\psi)$ are high. If both the values of $AS(\phi \Rightarrow \psi)$ and $AS(\psi \Rightarrow \phi)$ are relatively low, or $G(\phi \wedge \psi)$ is relatively low, we say that ϕ and ψ are negatively associated.
- **Conditional association rules**: Consider three concepts ϕ,ψ and ϕ'. If the difference $AS(\phi \wedge \phi' \Rightarrow \psi \wedge \phi') - AS(\phi \Rightarrow \psi)$ is high, we say that ϕ strongly supports ϕ under the condition ϕ'.

All these types of knowledge may find some practical applications. In general, one can study many different types of knowledge.

3.2. Technique levels

At the technique levels, we focus on algorithms and methods for extracting knowledge from data. Since the main stream of researches in data mining fall in these levels, we will not discuss them in detail.

A few points are perhaps worthy mentioning. Studies on technique levels should be as much application independent as possible. It should be realized that the existence of an algorithm does not necessarily imply that the discovered knowledge is meaningful and useful. Each algorithm should contain adjustable parameters that can be controlled at the application levels. Furthermore, it is desirable that those parameters can be related to criteria in the application levels.

The division between technique levels and application levels is not a clear cut. It may be desirable to develop specific data mining algorithms with respect to a particular domain of applications.

3.3. Application levels

The ultimately goal of data mining is to provide useful and meaningful knowledge. The application levels therefore should focus on the notions of "usefulness" and "meaningfulness" of discovered knowledge. Although these notions can not be discussed in total isolation with applications, one still can provide some general models. In the rest of the section, we provide a utility model in attempt to define "usefulness" and an explanation model to define "meaningfulness".

3.3.1. A utility model

The "usefulness" of discovered knowledge may be formally examined based on utility theory.[7] The theory is particularly relevant, if the discovered knowledge is to be used to assist decision making. We briefly discuss utility based association rules mining.[15, 26]

Suppose a department store wants to use association rules to improve sales and services. Association rules may be helpful for the tasks such as recommendation, stock management, and stock control. For an easy discussion, we re-express an association in the form $A \Rightarrow B$, where A and B are subsets of items. Using the knowledge given by an association, one may display items in A and B adjacent to each other to achieve implicit recommendation or increase the convenience of customers. Association rules can be used to design promotional packages containing several items. All those activities aim at achieving more profit or increase customer satisfaction. It becomes clear that the usefulness of an association rule may be measured in terms of its value.

Formally, the usefulness of an association can be measured in terms of a utility function on the set of items. Let $u : I \longrightarrow \Re^+$ be a utility function from the set of items I to the set of non-negative real $\Re^+$. The utility may be determined by more practical terms, such as sale price, cost and overhead. For a subset of items A, its utility may be defined through additivity, namely,

$$u(A) = \sum_{i \in A} u(i). \tag{9}$$

In general, utility functions is not additive. Depending on particular applications, utility functions may be interpreted in terms of profit, privacy, interesting, or uncertainty.

3.3.2. An explanation model

A piece of discovered knowledge is meaningful and trustful only if we have an explanation. An association does not immediately offer an explanation. One needs to find explanations regarding when, where, and why an association occurs.

Yao *et al.*[30] proposed an explanation oriented model for mining association based on the notion of conditional association (rules). With a conditional association, one can explicitly state the conditions under which the association occurs. Such conditions can provide additional semantics to a standard association. By adding time, place, and customer features or profiles as conditions, we can identify when, where and why an association occurs, respectively.

Conceptually, the proposed method consists of two parts and uses two data tables. A transaction data table is used to learn an association in the first step. An explanation table is used to construct an explanation of the association in the second step. In a wider context, explanation oriented model can be used to explain results from unsupervised learning. In the first step, an unsupervised learning algorithm is used to discover an piece of knowledge represented as a concept. In the second step, the extension of the concept is used as a set of positive instances, the rest as negative instances, and a supervised learning method is used to construct an explanation.

4. CONCLUSION

This paper is a first step towards the foundations of data mining. We focus on identifying important issues, rather than providing answers. While it is easy to demonstrate the needs for the study of foundations of data mining, it is difficult to decide what are the proper or correct approaches. As pointed out by Simpson,[21] the foundations of field often take time to develop and may change with the increasing understanding of the fields. The study on the foundations of data mining may make a significant contribution to the field.

REFERENCES

1. Berry, M.J.A. and Linoff, G.S. *Mastering Data Mining*, John Wiley & Sons, Inc., New York, 2000.
2. Boose, J. and Gaines, B. Foundations of knowledge acquisition, in: *The Foundations of Knowledge Acquisition*, Boose, J. and Gaines, B. (eds.) Academic Press, London, xi-xxi, 1990.
3. Chen, Z. The three dimensions of data mining foundation, *Proceedings of IEEE ICDM'02 Workshop on Foundation of Data Mining and Knowledge Discovery*, 119-124, 2002.
4. Demri, S. and Orlowska, E. Logical analysis of indiscernibility, in: *Incomplete Information: Rough Set Analysis*, Physica-Verlag, Heidelberg, 347-380, 1998.
5. Fayyad, U.M., Piatetsky-Shapiro, G. (eds.) *Advances in Knowledge Discovery and Data Mining*, AAAI Press, 1996
6. Frawley, W.J., Piatessky-Shapiro, F. and Matheus, C. Knowledge discovery in databases: an overview, in: *Knowledge Discovery in Databases*, Piatessky-Shapiro, F. and Frawley, W.J. (eds.), AAAI/MIT Press, Cambridge, 1-30, 1991.

7. French, S. *Decision Theory, an Introduction to the Mathematics of Rationality*, Ellis Horwood Limited, Chichester, 1986.
8. Gehrke, J. Report on the SIGKDD 2001 conference panel "New Research Directions in KDD", *SIGKDD Explorations*, **3**(2), 76-77.
9. Gunopulos, D. and Rastogi, R. Workshop report: 2000 ACM SIGMOD workshop on research issues in data mining and knowledge discovery, *SIGKDD Explorations*, **2**(1), 83-84.
10. Han, J. and Kamber, M. *Data Mining: Concept and Techniques*, Morgan Kaufmann Publisher, 2000.
11. Hand, D.J. Statistics and data mining, intersecting disciplies, *SIGKDD Explorations*, **1**(1), 16-19, 1999.
12. Lin, T.Y. Issues in modeling for data mining, *Proceedings of the International Conference on Computer Software and Applications*, 1152-1157, 2002.
13. Lin, T.Y. and Liau, C.J. (eds.) *Proceedings of the PAKDD'02 Workshop on Fundation of Data Mining, Communications of Institute of Information and Computing Machinery*, Vol. 5, No. 2, 101-106.
14. Lin, T.Y. and Ohsuga, S. (eds.) *Proceedings of IEEE ICDM'02 Workshop on Foundation of Data Mining and Knowledge Discovery*, 2002.
15. Lin, T.Y., Yao, Y.Y. and Louie, E., Value added association rules, *Proceedings of PAKDD'02*, 328-333, 2002.
16. Mannila, H., Theoretical frameworks for data mining, *SIGKDD Explorations*, **1**(2), 30-32, 2000.
17. Michalski, R.S., Carbonell, J.G., and Mitchell, T.M. (Eds.), *Machine Learning, an Artificial Intelligence Approach*, Morgan Kaufmann Publishers, Inv., Los Altos, California, 1983.
18. Pawlak, Z. *Rough Sets, Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers, Dordrecht, 1991.
19. Posner, M.I. (ed.) *Foundations of Cognitive Science*, The MIT Press, Cambridge, Massachusetts, 1989.
20. Simon, H.A. and Kaplan, C.A. Foundations of cognitive science, in: *Foundations of Cognitive Science*, Posner, M.I. (ed.), The MIT Press, Cambridge, Massachusetts, 1-47, 1989.
21. Simpson, S.G. What is foundations of mathematics? 1996, http://www.math.psu.edu/simpson/hierarchy.html
22. Smith, E.E. Concepts and induction, in: *Foundations of Cognitive Science*, Posner, M.I. (ed.), The MIT Press, Cambridge, Massachusetts, 501-526, 1989.
23. Sowa, J.F. *Conceptual Structures, Information Processing in Mind and Machine*, Addison-Wesley, Reading, Massachusetts, 1984.
24. Suzuki, E. Autonomous discovery of reliable exception rules, *Proceedings of KDD'97*, 259-262, 1997.
25. Van Mechelen, I., Hampton, J., Michalski, R.S. and Theuns, P. (eds.) *Categories and Concepts, Theoretical Views and Inductive Data Analysis*, Academic Press, New York, 1993.
26. Wang, K., Zhou, S. and Han, J. Profit mining: From patterns to actions. In: *Proceedings of EDBT*, 70-87, 2002.
27. Wille, R. Concept lattices and conceptual knowledge systems, *Computers Mathematics with Applications*, **23**, 493-515, 1992.
28. Xie, Y. and Raghavan, V.V. Probabilistic logic-based characterization of knowledge discovery in databases, *Proceedings of IEEE ICDM'02 Workshop on Foundation of Data Mining and Knowledge Discovery*, 107-112, 2002.
29. Yao, Y.Y. Modeling data mining with granular computing, *Proceedings of the 25th Annual International Computer Software and Applications Conference (COMPSAC 2001)*, 638-643, 2001.
30. Yao, Y.Y., Zhao, Y. and Maguire, R.B. Explanation oriented association mining using rough set theory, to appear in *Proceedings of RSFSGrC'03*, 2002.
31. Yao, Y.Y. and Zhong, N. An analysis of quantitative measures associated with rules, *Proceedings of PAKDD'99*, 479-488, 1999.
32. Zhong, N., Yao, Y.Y. and Ohsuga, S. Peculiarity oriented multi-database mining, *Proceedings of PKDD'99*, 136-146, 1999.