Explanation Oriented Association Mining Using Rough Set Theory

Y.Y. Yao, Y. Zhao, R.B. Maguire

Department of Computer Science, University of Regina
Regina, Saskatchewan, Canada S4S 0A2
E-mail: {yyao, yanzhao, rbm}@cs.uregina.ca

Abstract. This paper presents a new philosophical view and methodology for data mining. A framework of explanation oriented data mining is proposed and studied with respect to association mining. The notion of conditional associations is adopted, which explicitly expresses the conditions under which an association occurs. To illustrate the basic ideas, the theory of rough sets is used to construct explanations.

1 Introduction

In this paper, we view data mining as a specific field of study concerning theories, methodologies, and in particular, computer systems for exploring and analyzing a large amount of data [2]. A data mining system is designed with an objective to automatically discover, or to assist a human expert to discover, knowledge embedded in data [4]. This view allows us to examine data mining in a wide context with respect to revolutions, theories, and creativity in science [3]. The examination in turn may help us to see the limitations of current data mining research and the needs for a reconsideration of fundamental issues of the field.

In the development of many branches of science such as mathematics, physics, chemistry, and biology, the discovery of a natural phenomenon is only the first step. The important subsequent tasks for scientists are to build a theory accounting for the phenomenon and to provide justifications, interpretations, and explanations of the theory. The interpretations and explanations enhance our understanding of the phenomenon and guide us to make rational decisions.

The general observation sheds new light on the study of data mining, if one draws analogy of human centered scientific discovery tasks and data mining tasks. Current research in data mining focuses mainly on the task of discovering a natural phenomenon, represented as patterns, rules, or clusters in data. Little attention is paid to the tasks of constructing theories and models that account for a discovered phenomenon. In other words, researchers concentrate on the task of detecting the *existence* of a pattern, and have not moved a step further to the task of searching for the underlying *reasons* that explain the existence of the pattern. Unless we take another step to the explanation task, effectiveness of data mining systems would be limited.

The main objective of this paper is to propose and evaluate a framework of explanation oriented data mining. On top of the traditional tasks of pattern

discovery and evaluation, we add the tasks of explanation construction and evaluation. Furthermore, we adopt supervised learning methods to search for and evaluate plausible explanations.

The contribution of the paper is *not* a new algorithm, but a new *philosophical view* and a new *methodology*, namely, data mining must deal with both discovery and explanation. One may argue that many researchers have in fact implicitly considered the issues we address here, and hence question the significance of the approach. If one takes another look at the whole picture in the wide context presented earlier, it is easy to come up with another conclusion. The recognition and identification, as well a clearly description, of the new philosophical view, the explicit separation of explanation construction and evaluation from other tasks, and a general framework unifying all these notions, may have a fundamental contribution to the development of the field.

2 A Framework of Explanation Oriented Data Mining

The notion of information tables is used as a formal model to represent and interpret various notions of data mining, such as data, patterns and rules [15]. A framework of explanation oriented data mining is presented and compared with related studies.

2.1 Basic notions of data mining in information tables

An information table is defined as a system:

$$S = (U, At, \mathcal{L}, \{V_a | a \in At\}, \{f_a | a \in At\}),$$

where U is a finite nonempty set of objects called universe, At is a finite nonempty set of attributes, $\mathcal{L}$ is a language, V_a is a nonempty set of values for each $a \in At$, and $f_a : U \to V_a$ is an information function mapping an object to a value in V_a. In the language $\mathcal{L}$, an atomic formula ϕ is given by $a = v$, where $a \in At$ and $v \in V_a$. All formulas of $\mathcal{L}$ are defined recursively by logical connectives $\neg$, $\wedge$, $\vee$, $\to$, and $\leftrightarrow$.

For an atomic formula $a = v$, an object x in U satisfies $a = v$ if $I_a(x) = v$. The satisfiability of all formulas can be easily established. Given a formula ϕ, we associate with it a set $m(\phi) = \{x \in U \mid x \text{ satisfies } \phi\}$ called the meaning of ϕ in S. In this way, we define a pair $(\phi, m(\phi))$, which represents the basic level knowledge in S. The pair $(\phi, m(\phi))$ is commonly known as a formal concept with ϕ as its intension and $m(\phi)$ as its extension [15].

The next level of knowledge can be summarized by relationships between concepts. Given two concepts $(\phi, m(\phi))$ and $(\psi, m(\psi))$, a relationship between them is written in a general form through their intensions as $\phi \circ \psi$. The symbol $\circ$ can be interpreted in many ways [16]. For example, $\circ$ may be interpreted as one-way associations denoted by $\Rightarrow$, two-way associations denoted by $\Leftrightarrow$, and similarity denoted by $\approx$.

Statistical information about two concepts are given by the contingency table in terms of their extensions:

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

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

In the table, $|\cdot|$ denotes the cardinality of a set, and $|U| = n$. Relationships of concepts can be defined, identified, classified, and interpreted using the information in the contingency table. For example, we have [15, 16]:

$$support(\phi \wedge \psi) = \frac{|m(\phi \wedge \psi)|}{|U|} = \frac{a}{n}, \tag{1}$$

$$similarity(\phi \approx \psi) = \frac{|m(\phi \wedge \psi)|}{|m(\phi \vee \psi)|} = \frac{a}{a+b+c}, \tag{2}$$

$$confidence(\phi \Rightarrow \psi) = \frac{|m(\phi \wedge \psi)|}{|m(\phi)|} = \frac{a}{a+b}, \tag{3}$$

$$coverage(\phi \Rightarrow \psi) = \frac{|m(\phi \wedge \psi)|}{|m(\psi)|} = \frac{a}{a+c}. \tag{4}$$

A task of data mining is to discover useful relationships between concepts based on these quantities. For example, in association rule mining, one is interested in finding all one-way associations $\phi \Rightarrow \psi$ whose support $support(\phi \wedge \psi)$ and confidence $confidence(\phi \Rightarrow \psi)$ are above certain threshold values.

2.2 Explanation oriented data mining

A data mining process typically consists of the following steps: data cleaning, data integration, data selection, data transformation, pattern discovery and evaluation, and presentation [5]. Conceptually, explanation oriented data mining inserts an additional step called explanation construction and evaluation into the data mining process. Its main function is to search for the underlying reasons that explain the existence of a discovered pattern. The basic idea of explanation oriented data mining is presented in this section with respect to association mining.

Suppose we have discovered an association from a transaction database by using the well-known Apriori algorithm [1]. The identification of an association is only the first step. We need to understand the meaning as well as the implications of the association. The previous discussed measures provide some useful information, but offer no explanation regarding *why* the association does exist. In many situations, such an explanation is needed, if one wants to justify any decision made based on the discovered pattern. One can be easily convinced that plausible explanations can not be obtained from the original transaction databases. In other words, we need to find explanations from other data source or related domain specific information.

Explanations for an association may be obtained in several ways. If additional information about customers is available, we may search for an explanation for

the association based on customer features. If the time of every transaction is available, we may explain the association in terms of transaction time. If products information is available, we may explanation the association through product characteristics. Each of these explanations may be useful to different types of users of a data mining system.

A method of explanation construction and evaluation is given below. A transaction database can be easily expressed as a binary information table called transaction table. Let ϕ denote a particular association in a set U of transactions.

1. We introduce a binary attribute named **association**. Given a transaction $x \in U$, its value on **association** is 1 if it satisfies ϕ in the original transaction table, otherwise its value is 0.
2. We select a set E of attributes related to possible explanations of association. For example, we can choose a set of attributes of customers who made the transactions.
3. We construct an information table by using the attributes obtained from (1) and (2). The new table is called an explanation table.
4. By treating **association** as the target class, we can apply any supervised machine learning method to derive classification rules of the form:

$$c \Rightarrow \textbf{association} = 1,$$

   which corresponds to the conditional association $\phi \mid c$. The condition c is a formula in the explanation table, which clearly states the condition c under which the association ϕ occurs.
5. We evaluate conditional associations based on previously discussed statistical measures.

With a conditional association, one explicitly states the conditions under which the association occurs. If conditional associations are constructed properly, they provide explanation for the original association.

The basic idea of explanation oriented data mining can be applied to explain results from any unsupervised learning method. Suppose one applies a unsupervised clustering algorithm to a data set. For a particular cluster, we can construct an explanation table with elements from the cluster as positive instances. A supervised machine learning algorithm can then be used to construct explanations for that cluster.

It should be noted that the above procedure for explanation construction and evaluation is only one of the possible solutions. The effectiveness of the procedure needs to be evaluated experimentally on real world data. An advantage of the procedure is that it only uses existing data mining and machine learning algorithms, and hence can be easily added to any existing data mining systems. Moreover, it provides another example to demonstrate that one can apply existing data mining technologies to obtain totally new results. Future data mining research needs to pay more attention to how to apply existing algorithms effectively, in addition to the study of new algorithms.

2.3 Comparisons with existing studies

Two specific research directions of data mining are closely related to explanation oriented data mining. They are pattern evaluation and multi-database mining. A comparative study of those topics will put explanation oriented data mining into a proper perspective.

A discovered piece of knowledge is considered to be interesting if it is novel (new), potentially useful, understandable, actionable, or profitable. Many proposals have been made to precisely quantify such an intuitive notion of interestingness. From extensive existing studies, one can identify roughly two classes of approaches, the statistics centered approaches and the semantics centered approaches. They lead to different philosophies in designing data mining solutions to real world problems.

Statistics centered approaches focus on the statistical characteristics of discovered patterns [16]. They are user, application and domain independent. A pattern is deemed interesting if it has certain statistical properties. Different classes of rules can be identified based on statistical characteristics [17], such as peculiarity rules (low support and high confidence), exception rules (low support and high confidence, but complement to other high support and high confidence rules), and outlier patterns (far away from the statistical mean). Although statistical information provides an effective indicator of the nature of a pattern, its usefulness is limited.

Semantics centered approaches are application and user dependent. In addition to statistical information, one incorporates other domain specific knowledge such as user interest, utility, value, profit, actionability, and so on. In *constraint based mining*, the user provides restraints that specify the type of knowledge, ranges of a measure or subset of database to be mined [6]. In *utility/profit based mining*, a profit is associated with each item and the usefulness of discovered patterns is determined by their profits [7, 13]. In *actionable data mining*, a pattern is considered to be useful only if it leads to profitable actions [8, 9].

Both statistical and semantical approaches of pattern evaluation focus on different features of patterns. Their main function is to filter out patterns that are not worthy further investigation. They only examine and summarize *characteristics* of the discovered patterns, and do not provide *explanation* for the existence of the pattern. On the other hand, measures and methods used in pattern evaluation can be applied to explanation evaluation.

In the framework of explanation oriented data mining, we explicitly use two sets of data. It is tempting to conclude that the proposed framework is simply a multi-database mining model. This conclusion on the superficial similarity is simply not valid. Existing studies on multi-database mining mainly focus on pattern discovery rather than explanation construction. Patterns discovered in multi-database mining also need to be explained. On the other hand, multiple data sources are used for quite different purpose in explanation oriented data mining. In general, explanation construction may be carried out by using domain knowledge without explicitly using a second data source.

3 Construction of Explanation Using Rough Set Theory

The theory of rough sets has been used as a basis for the design and implementation of many supervised machine learning algorithms [10, 11]. There are two types of learning algorithms based on rough set theory, namely, attribute oriented induction and granule (i.e., attribute-value pair) oriented induction [14]. One can immediately apply rough set based methods to the explanation table to construct explanation.

A key notion of attribute oriented induction is reduct. With respect to the attribute **association**, a reduct is a set of attributes that individually necessary and jointly sufficient to define **association**. In other words, we need each attribute in a reduct to explain the **association**, and all attributes in a reduct together are sufficient to explain the **association**.

Many algorithms have been proposed for computing a reduct. Let E be the set of attributes in the explanation table. The following is an outline of algorithm for finding a reduct:

1. Let $R = \emptyset$ and $S = E$,
2. If R is a reduct or $S = \emptyset$, exit,
3. Select an attribute a from S based on a criterion,
4. Add a to R and remove a from S,
5. Go back to step 2.

The criterion in step 3 can be defined in terms of lower and upper approximations. It can also be defined based on user preference of attributes. The condition in step 2 can also be modified so that the procedure would stop as soon as a set of attributes that provide reasonable explanations is found. From a reduct R, all combinations of attribute values in R will be used as conditions c in the conditional association $\phi \mid c$. Those conditional associations that meet certain criteria will be presented as potential explanations.

Alternatively, we can apply granule oriented induction. In this case, we focus on a single explanation each time. An outline of such an algorithm is given by:

1. Let $c = \mathbf{1}$,
2. Let $a = v$ denote an atomic formula maximizing $support(\phi \mid (a = v))$,
3. Add $a = v$ to c by conjunction,
4. If $support(\phi \mid c) > support(\phi)$, exit; otherwise, goto step 2.

In the algorithm, the support of conditional association is given by:

$$support(\phi \mid c) = \frac{|m(\mathbf{association} = 1 \wedge c)|}{|m(c)|}. \tag{5}$$

The condition in step 4 suggests that under condition c the association is more pronounced, and thus may provide a plausible explanation. Other evaluation criteria can also be used in step 4.

We have carried out a real experiment to evaluate the effectiveness of explanation oriented mining. We generated a Web site consisting of pages on several

topics, such as Business, Finance, and so on. We built a user profile database by collecting relevant information believed to be related to user browsing pattern. By applying the Apriori algorithm on Web log files, we found an association $\phi = Business \wedge Finance$ with support of 5.71%. To construct explanation for this association, we used the Rosetta system [12], a rough set based tool kit, to learn conditional association from the explanation table. The following results were obtained:

- reduct length is one:
  - $support(\phi \mid age = [30, 49]) = 9.1\%$;
  - $support(\phi \mid gender = female) = 8.0\%$;
  - $support(\phi \mid occupation = student) = 6.7\%$;
  - $support(\phi \mid occupation = employee) = 5.7\%$.
- reduct length is two:
  - $support(\phi \mid gender = male \wedge age = [30, 49]) = 13.0\%$;
  - $support(\phi \mid occupation = employee \wedge age = [20, 29]) = 6.0\%$.

We got a set of conditional associations with the support value being higher than that of the unconditional association. The set of conditions provides some interesting explanations of the browsing pattern. The results clearly show that age, gender and occupation influence the users' browsing behavior. Such knowledge is useful to a Web site designer.

4   Conclusion

A new philosophical view and methodology called explanation oriented data mining is introduced. It is argued that the effectiveness of current data mining systems is unnecessarily limited by a lack of explanation of discovered knowledge. In order to resolve this problem, we suggest to insert another step, namely, explanation construction and evaluation, to the commonly accepted data mining precess.

A framework of explanation oriented mining is proposed. It is recognized that explanations may not be found using the original data table. Additional information is collected to form an explanation table. The results of current data mining systems can be used to define a classification problem in the explanation table. Consequently, any standard machine learning algorithm can be used to construct plausible explanations.

Association mining is used to illustrate the main idea of explanation oriented data mining. Explanations are coded as conditions in conditional associations. Rough set based machine learning algorithms are used to search for such conditions.

References

1. Agrawal, R., Imielinski, T. and Swami, A. Mining association rules between sets of items in large databases, *Proceedings of SIGMOD*, 207-216. 1993.

2. Berry, M.J.A. and Linoff, G.S. *Mastering Data Mining*, John Wiley & Sons, Inc., New York, 2000.

3. Bohm, D. and Peat, F.D. *Science, Order, and Creativity*, 2nd Edition, Rooutledge, London, 2000.

4. Fayyad, U.M., Piatetsky-Shapiro, G. (eds.) *Advances in Knowledge Discovery and Data Mining*, AAAI Press, 1996.

5. Han, J. and Kamber, M. *Data Mining: Concept and Techniques*, Morgan Kaufmann Publisher, 2000.

6. Han, J., Lakshmanan, L. and Ng, T. Constraint-based, multidimensional data mining, *Computer magzine*, 1999.

7. Lin, T.Y., Yao, Y.Y. and Louie, E. Mining value added association rules, *Proceedings of PAKDD'02*,

8. Ling, C., Chen, T., Yang, Q. and Chen, J. Mining optimal actions for intelligent CRM, *Proceedings of ICDM*, 2002.

9. Liu, B., Hsu, W. and Ma, Y. Identifying non-actionable association rules, *Proceedings of KDD*, 329-334, 2001.

10. Pawlak, Z. *Rough Sets: Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers, Dordrecht, 1991.

11. Polkowski, L. and Skowron, A. (Eds.) *Rough Sets in Knowledge Discovery I, II*, Physica-Verlag, Heidelberg, 1998.

12. Rosetta, a rough set toolkit for analyzing data, http://www.idi.ntnu.no/ãleks/rosetta

13. Wang, K., Zhou, S. and Han, J. Profit mining: from patterns to actions, *Proceedings of EDBT*, 70-87, 2002.

14. Yao, J.T. and Yao, Y.Y. Induction of classification rules by granular computing, *Proceedings of RSCTC'02*, 331-338, 2002.

15. 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.

16. Yao, Y.Y. and Zhong, N. An analysis of quantitative measures associated with rules, *Proceedings of PAKDD*, 479-488, 1999.

17. Zhong, N., Yao, Y.Y., Ohshima, M. and Ohsuga, S. Interestingness, peculiarity, and multi-database mining, *Proceedings of ICDM*, 566-573, 2001.