Finding Interesting Rules Exploiting Rough Memberships

Lipika Dey[1], Amir Ahmad[2], and Sachin Kumar[1]

[1] Department of Mathematics, Indian Institute of Technology, Delhi, Hauz Khas, New Delhi - 110 016, India
lipika@maths.iitd.ac.in

[2] Solid State Physics Laboratory, Timarpur, Delhi, India - 110 054
amir_ahmedsspl@ssplnet.org

Abstract. In this paper we propose a method to identify significant attributes which aid in good classification, as well as introduce certain degrees of roughness. Exception rules are formed with these attributes using the fact that exceptional elements have high rough memberships to more than one distinct class.

1 Introduction

Knowledge discovery refers to the generic task of discovering interesting, hidden, potentially useful patterns embedded in the data. Interestingness measure may be user or application dependent. In this work, we consider only rare or exceptional patterns as interesting. We propose a rough set-based method for mining rare and unexpected associations from a database in a guided fashion.

It is obvious that exceptions arise when the roughness of a set is high. An exceptional element has a high rough membership value to classes other than its defined class. Given a subset of attributes, it is easy to find the degree of roughness and this can be used to identify the exceptions. However, finding the ideal subset of attributes is a difficult task. We observe that though significant attributes which provide good classification are usually the ones which introduce low degrees of roughness, insignificant attributes do not provide correct classification knowledge. Thus our aim is to find rules based on significant attributes which also introduce roughness into a database. We propose methods to find two categories of exception rules. The first set identifies contradictions to usual classificatory knowledge. The second set of rules are compound in nature identifying combinations of appropriate features that lead to exception formation. This work is related to the classification scheme reported in [2]. The rest of the paper is organized as follows. In section 2 we provide a short overview of related work. Sections 3, 4 and 5 describe the proposed techniques for finding exceptions and interesting rules. We have explained the method through examples.

S.K. Pal et al. (Eds.): PReMI 2005, LNCS 3776, pp. 732–737, 2005.

2 Rule Interestingness Measures – A Brief Survey

Interestingness measures proposed in literature are of two kinds - objective and subjective [5]. Among objective measures, Piatetsky-Shapiro's rule interestingness measure [8] quantifies the correlation between attributes in a simple classification rule. [11] proposed the J-measure which is the average information content of a probabilistic classification rule. Agrawal and Srikant's item set measures [1] identify frequently occurring associations in large databases. [3] proposes two objective rule-interestingness measures based on a one-attribute-at-a-time basis. In [10] interestingness is defined as the extent to which a soft belief is changed as a result of encountering new evidence i.e. discovered knowledge. In [14], a new interestingness measure, peculiarity, has been defined as the extent to which one data object differs from other similar data objects. [6] states how learned rules can be analyzed for interestingness using probabilistic measures. Objective measures do not take user perspective into account. [7] proposes a subjective approach for selecting interesting rules based on the notion of general impressions specified by users. [9] describes a genetic algorithm designed for discovering interesting fuzzy rules based on general impressions. The subjective approach assumes user is aware of common associations, while it may not be really so. [4] suggests use of objective interestingness measures to filter truly interesting rules and then use subjectiveness. Our approach is also intermediary.

3 Finding Significant Attributes

In [2] it has been shown that for a significant attribute with good classificatory power, there is a strong possibility that elements with complementary sets of values for this attribute, belong to complementary sets of classes. Alternatively, given that the class decisions for two sets of elements are different, the significant attribute values for these two sets of elements are usually different. These measures are computed as follows.

Let U be a collection of pre-classified elements belonging to m different classes, described by attributes A_1, A_2, $\cdots$, A_g. Let C represent any class from this set. Let J represent the set of attribute values that an attribute A_i can take. Let W denote a proper subset of J. We introduce $\alpha(A_i)$ to capture the cumulative effect of all possible values of A_i and their positive as well as negative associations to class decisions.

Definition 1. *$\alpha(A_i)$ is called the discriminating power of A_i.*

$$\alpha(A_i) = \Sigma_C(P_i^C(W') + P_i^{\sim C}(\sim W')),$$

where $W' \subset J$ such that $P_i^C(W') + P_i^{\sim C}(\sim W') > (P_i^C(W) + P_i^{\sim C}(\sim W) \forall W \neq W'$
$P_i^C(W')$ = Probability that elements of U with class label C have their ith attribute value belonging to class W',
$P_i^{\sim C}(\sim W')$ = Probability that elements of classes other than C, do not have a value in W' as their ith attribute value.

W' is called the support set of C with respect to attribute A_i. Values in W' have maximal positive association to class C and maximal negative association to other classes. Both the probabilities can be computed from frequency counts. $\beta(A_i)$ is similarly computed as the cumulative effect of positive and negative class correlations to each value of attribute A_i. Details of these computations are available in [2].

Definition 2. *The significance of an attribute is denoted by $\sigma(A_i)$ and is the mean of $\alpha(A_i)$ and $\beta(A_i)$. A high value of $\sigma(A_i)$ denotes an attribute with more classificatory powers.*

Examples from heart database [15] are produced below to show how this measure can find contradictions. Elements of heart database are classified as non-heart patients and heart patients. A common user belief is *"Serum Cholesterol is High implies patient has heart disease."* However, using our method *serum cholesterol* is found to have $\alpha() = 0.132$ and $\beta() = 0.162$, which are low compared to the most significant attributes like *thal* or *number of major vessels colored by fluorosopy* (0-3). Hence, this is a contradiction. On verification with the database it is found that 38% of people with low cholesterol do have heart disease, and 46% of the people with high cholesterol actually are classified as non-heart patients. The support sets for non-heart and heart patients for *thal* are {3} and {6,7} respectively, and for the other attribute these are {0} and {1,2,3} respectively. These are used later to find interesting rules.

4 Finding Exceptions and Interesting Rules Using Support Sets and Rough Memberships

Significant attributes in a database can correctly classify elements of a class based on these attribute values only. However, exceptions to a class are elements which cannot be predicted correctly even by the significant attributes. Support set values have strong positive correlations to the class decision and negative correlation to the complement of the class. Elements of a class which do not have values for significant attributes in the support set can therefore be termed as exceptions. Exceptions can be found from a database using rough memberships of elements to various classes in conjunction with their support sets. Exception rules can thereafter be formed as specified by the following definitions.

Definition 3. *Let $S_A(C)$ define the support set of class C with respect to attribute A. $[S^C]_A$ is defined as the collection of all elements x whose values for attribute A lie in the support set of class C.*

Definition 4. *Given a subset of attributes A, rough membership of an element x to a class C is computed as $\mu_C^A(x) = \frac{|[S^C]_A \bigcap C|}{|[S^C]_A|}$.*

Definition 5. *Let A denote a set of significant attributes. An element x is said to be an exception to class C, if for a given set of significant attributes A, $\mu_{C'}^A(x) > \mu_C^A(x)$, for some class $C' \neq C$.*
An exception rule is formed as $S_A(C) \implies C'$.

We now present some sample exception rules extracted from the German credit card [15] and heart databases, using this approach. German credit card database has good and bad creditors. The most significant attribute identified is *Status of existing checking account* with values $A11 :< 0DM, A12 : 0 <= \cdots < 200DM, A13 : ... >= 200DM$ and $A14$: *no checking account*. $A13$ and $A14$ belong to the support set of good creditors. $A11$ and $A12$ lie in the support set of bad creditors. The second most significant attribute for the data set is *credit amount*, which being numeric, has been discretized to HIGH, LOW and VERY LOW. HIGH is in support set of bad creditor and the other two are in the support set of good creditor. With an expected minimum support of 2%, and confidence 10%, some interesting exception rules extracted from these two databases using this approach are shown in Table 1.

Table 1. Exception rules for German Credit card and Heart databases

Antecedent	Class	Support	Confidence
Status of existing checking account = A13 or A14 and *Credit amount* VERY LOW	bad creditor	4.8%	11%
Status of existing checking account = A11 or A12 and *Credit amount* = HIGH	good creditor	3.6%	41%
thal = 6 or 7 *number of major blood vessels colored by fluoroscopy* $= 1, 2 or 3$	not heart patient	2.2%	9%

5 Identifying Attribute Combinations That Cause Exceptions

While significant attributes help in classification, attributes which introduce roughness into a data set are the ones which cause exceptions. In this section we define *roughness coefficient* for each attribute. This along with the significance value enable us to formulate exception rules.

Definition 6. *Let $s \in S_A(C)$. Let P and P' denote the number of elements which have value s and belong to class C and C' respectively. Then, $r_s(A) = \frac{|P'|}{|P|}$ denotes the degree of roughness introduced by the value s. The* roughness coefficient *of the attribute A is computed as mean of r_s for all values s of A and is given by $\rho(A) = 1/d * \Sigma_s(r_s(A))$, where attribute A has d values.*

Following is a detailed scheme to find interesting rules from a pre-classified database, using roughness coefficient of significant attributes.

Step 1: Find the most significant attribute of database.
Step 2: Transform the given database by fusing the most significant attribute and the class decision to give new class decisions. Let A be the most significant

attribute of original data set, which has been fused with class column. If A had n distinct values and there were m distinct classes, then there will be mn new class decisions. Let $v \in S_A(C)$. Then vC and vC' are two new decision classes. Our aim is to find support set for vC' in the new database, since this identifies feature values that in combination with v changes the potential class of an element from C to C'. To do this we find the attributes with highest mean of significance and roughness coefficients with respect to the new fused decision.
Step 3: Find $\sigma(A_i)$ and the support sets for all remaining attributes A_i with respect to the new decision classes.
Step 4: Find roughness coefficients $\rho(A_i)$ for all attributes.
Step 5: Arrange attributes in decreasing order of mean value of $\sigma(A_i)$ and $\rho(A_i)$.
Step 6: Form compound exception rules with the new-found attributes and the most significant attributes of the original data set as follows:
$v \in S_A(C)$. Let s be a value of a top-ranking attribute A_i in fused data set, which belonged to the support set of class C' in original data set. If s belongs to support set of class vC' in fused data set, then an exception rule of the form $sv \implies C'$ is formed.
Step 7: Repeat steps 2-7 till no more rules satisfying user-given support can be extracted.

Table 2 cites some compound interesting exception rules identified for the German Credit card and heart databases using the above approach.

Table 2. Compound exception rules for German Credit card and Heart databases

Antecedent	Class	Support	Comment
Status of existing checking account = A13 or A14 and *Savings accountbond* $< 100DM$	bad creditor	3.6%	Low savings changes class from good to bad
Status of existing checking account = A11 or A12 and *Property* = real estate	good creditor	10.3%	Real estate changes class from bad to good
thal = 3 and *maximum heart rate* is LOW	heart patient	5.18%	Bad heart Rate causes heart disease
thal = 6 or 7 and *resting blood pressure* is LOW	not heart patient	7.4%	Low blood pressure changes class

6 Conclusion

Interestingness of a rule depends on the perspective of the user. In this paper, we have presented an approach which can help the user in finding contradictions to popular beliefs and also discover exceptions in data sets. The approach is based on rough memberships of elements to different classes and can yield surprising associations. The method has been shown to be useful in identifying interesting exceptions in financial and medical databases.

References

1. Agrawal, R., Srikant, R.,: Fast Algorithms for Mining Association Rules. In Proceedings of the 20th International Conference on Very Large databases, Santiago Chile (1994) 487-499.
2. Ahmad, A., and Dey, L.,: A Feature Selection Technique for Classificatory Analysis. Pattern Recognition Letters, Vol. 26, (2005) 43-56.
3. Freitas, A.A.,: On Objective Measures of Rule Surprisingness. In J.Zytkow and M.Quafou, editors, Proceedings of the Second European Conference on the Principles of Data Mining and Knowledge Discovery ,Nantes France (1998) 1-9.
4. Freitas, A.A.,: Data Mining and Knowledge Discovery with Evolutionary Algorithms. Springer-Verlag (2002), ISBN: 3-540-43331-7.
5. Hilderman, R., Hamilton, H.,: Knowledge Discovery and Interestingness Measures: A Survey. Technical Report CS 99-04, Department of Computer Science, University of Regina (1999).
6. Liu, B., and Hsu, W.,: Post-Analysis of Learned Rules. In Proceedings of the Thirteenth National Conference on Artificial Intelligence (1996) 828–834.
7. Liu, B., Hsu, W., Chen, S.,: Using General Impressions to Analyze Discovered Classification Rules, Proceedings of the Third International Conference on Knowledge Discovery and Data Mining (KDD-97, full paper), Newport Beach, California, USA (1997)pp. 31-36.
8. Piatetsky-Shapiro, G.,: Discovery Analysis and Presentation of Strong Rules. In Knowledge Discovery in Databases ,AAAI/MIT Press (1991) 229-248.
9. Romao, W., Freitas, A.A., Pacheco, P.C.S.Genetic. Algorithm for Discovering Interesting Fuzzy Prediction Rules: applications to science and technology data In WB Langdon and E Cantu-Paz et al, editors, Proceedings of Genetic and Evolutionary Computation Conference (GECCO-2002), San Francisco, CA, USA, (2002) 1188-1195.
10. Silberschatz, Avi., Tuzhilin, Alxender.,: What Makes Patterns Interesting in Knowledge Discovery Systems. IEEE Transactions on Knowledge and Data , Volume 8 (1996) 970-974.
11. Smyth, P., and Goodman, R.M.,: Rule Induction Using Information Theory. In Knowledge Discovery in Databases, AAAI/MIT Press (1991) 159-176.
12. Sahar, S.,: On Incorporating Subjective Interestingness Into the Mining Process. IEEE International Conference on Data Mining (2002).
13. Wang, K., Jiang, Y. and Lakshman, L.V.,S.,: Mining Unexpected Rules by Pushing User Dynamic. Conference on Knowledge Discovery in Data , Proceedings of the ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining ,ACM Press New York USA (2003) 246-255.
14. Zhong, N., Yao, Y.Y., Oshuga S.,: Peculiarity-Oriented Multi-Database Mining. In J.Zytkow and J.Rauch (editors), Proceedings of the Third European Conference on the Principles of Data Mining and Knowledge Discovery, Czech republic (1999) 136-146.
15. www.liacc.up.pt/ML/statlog/dataset.html.