High confidence association rules for medical diagnosis

Dragan Gamberger (1), Nada Lavrač (2), Viktor Jovanoski (2)

(1) Rudjer Bošković Institute
Bijenička 54 10000 Zagreb, Croatia
dragan.gamberger@irb.hr
(2) Jožef Stefan Institute
Jamova 39 1000 Ljubljana, Slovenia
nada.lavrac@ijs.si, viktor.jovanoski@ijs.si

Abstract

This paper elaborates a simple and general decision model based on the so-called confirmation rules. Confirmation rules are generated separately for each diagnostic class so that selected rules cover (and should hence be able to reliably predict) a significant number of cases of the target class. At the same time, a confirmation rule should not cover the cases of non-target diagnostic classes, and when used for prediction it should exclude the possibility of classifying any of the non-target cases into the target class. In this work we have used and tested the association approach for rule generation, accepting only extremely high confidence rules with reasonable support level as potentially good confirmation rules. Experimental results in the problem of coronary artery disease diagnosis illustrate the approach.

1 Introduction

The general problem of the induction of reliable diagnostic rules is hard because theoretically no induction process by itself can guarantee the correctness of induced hypotheses. In practical situations the problem is even more difficult due to unreliable diagnostic tests and the presence of noise in training examples. This may result in hypotheses with unsatisfactory prediction accuracy which are too unreliable for critical medical applications.

A solution of the problem of reliable diagnostic rules is the construction of either very sensitive or very specific rules instead of rules with a high overall prediction accuracy. This solution originates from the observation that a false negative classification (for example, a patient with cancer classified as healthy) has not to be as dangerous as a false positive classification (a healthy patient classified as ill) [9]. An alternative approach to the construction of reliable diagnostic rules is the construction of redundant rules, which is known to be appropriate for achieving reliable predictions. It was experimentally demonstrated that the prediction accuracy can be improved by combining different classifiers for the same domain [2, 13]. In most cases classifiers are combined by voting to form a compound classifier. Different classifiers can be obtained either by the application of different learning algorithms on the same training set or by the same learning algorithm on different training (sub)sets. The later approach is used in the well-known *bagging* and *boosting* approaches that employ redundancy to achieve better classification accuracy [3, 4, 11]. For critical applications the prediction accuracy of compound classifiers can be further increased if, instead of voting, the consensus of classifiers' answers is requested. The main disadvantage of compound classifiers is that the independence of classifiers can not be guaranteed, which means that also the prediction reliability of such classifiers can not be ensured in all situations.

This paper elaborates an approach to reliable medical diagnosis based on the confirmation rules decision model. The main difference to other standard decision models is that this method does not aim at giving a decisive answer in every situation. In this sense the approach follows the concept of *reliable, probably almost always useful learning* defined in [12]. This means that in the case of a two-class problem, three different possible predictions are considered: class positive, class negative, and answer not possible. By considering three possible answers it can be ensured that the method gives only reliable answers; this is the main advantage of the approach. A disadvantage of this method are indecisive answers, whose amount has to be kept as low as possible.

An application of the confirmation rules decision model in the domain of coronary artery disease diagnosis is presented in [6]. This paper upgrades the results in the following sense:

a) An association rule learning algorithm was used for constructing confirmation rules.

b) Cross-validation was used for evaluating the expected prediction results.

c) Experiments with and without noise detection and elimination in preprocessing were performed.

The confirmation rules decision model is described in Section 2 while Section 3 presents the association search based algorithm for the construction of confirmation rules. Experimental results obtained in the problem of coronary artery disease diagnosis are described and analised in Section 4.

2 Confirmation rules

In the decision model based on confirmation rules, every diagnostic class is treated separately as the target class. For a given target class a rule is a conjunction of logical tests (literals). Confirmation rules have a similar form as, for example, association rules [1] and if-then rules [10] as induced

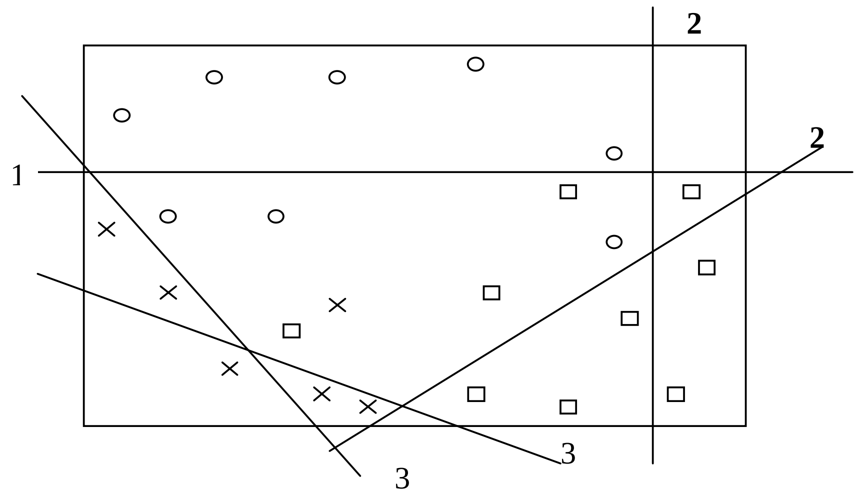

Figure 1: The confirmation rule concept illustrated on a three-class problem. Surfaces between the individual lines and the borders represent the areas covered by the individual confirmation rules.

by the AQ15 learning system. The main difference with association rules is that confirmation rules have only the class assignment in the conclusion of a rule whereas a conclusion of an association rule is a conjunction of arbitrary attribute values. On the other hand, the main difference compared to AQ generated rules is that every *complex* (conjunction) of an AQ rule is in the context of confirmation rules treated as a separate and independent rule.

The concept of confirmation rules is graphically presented in Figure 1. Confirmation rules have the following properties: a confirmation rule has to cover a significant number of cases of the target class and at the same time a confirmation rule should not cover cases of non-target diagnostic classes. The consequence is that every confirmation rule can be used independently of other confirmation rules or in combination with any subset of other confirmation rules. For a given unclassified case, the following outcomes are possible:

a) If no confirmation rule fires for the case, class prediction is indecisive (the case is not classified).

b) If a single confirmation rule fires for the case, class prediction is determined by this rule.

c) If two or more confirmation rules of the same class fire for the case, this class is predicted with increased reliability.

d) If two or more confirmation rules fire for the case and at least two of these rules are for different classes, class prediction is indecisive.

This indicates that the confirmation rules do not give a decisive prediction in every situation (cases (a) and (d)), and that a prediction of increased reliability can be achieved (case (c)).

3 Confirmation rule construction by association search

Prediction quality of the confirmation rule based decision model depends on the quality of the induced confirmation rules. Exhaustive search, because of its time complexity, is not an appropriate solution for the construction of high quality confirmation rules. Recently, construction of association rules was shown to be an effective approach for rule induction in large databases [1]. Its main drawback is that generally many rules are suggested among which it is difficult to select the best ones. The idea of this work is to use the association search in order to generate rules and to select only those which satisfy the properties of confirmation rules.

Algorithm 1 presents an approach to the generation of a set of confirmation rules for a selected target class based on association search. The quality of the generated confirmation rules can be easily controlled by the selection of the acceptance parameter *min_support* (step 1). The *min_support* parameter indirectly determines also the total number of generated rules. If this number is large, the *max_rules* parameter determines the maximal number of rules that will be included into the output set. It must be noticed that Algorithm 1 is not a simple selection of *max_rules* number of rules with maximal support. Experiments showed that in the case of such a simple approach, the constructed rules are very similar, they cover almost the same set of target class examples, and practically the same attribute subset is used in their logical tests. The consequence is that such a rule set covers a relative small number of unseen examples resulting in many indecisive predictions. Also, the increased reliability, due to positive predictions by more than one confirmation rule, is not very trustworthy because it is based on the same/similar attribute values. To overcome this problem, in Algorithm 1, a separate counter $c(e)$ for every target class example $e \in P$ is introduced. In step 4 these counters are preset to value 1. After that, the loop (steps 5–10) is started *max_rules* times and in each iteration one confirmation rule is added to the output set. Rule selection is based on the maximization of the sum $\sum 1/c(e)$ for the target class examples that are covered by the rule. In the first iteration when all $c(e)$ values are equal 1, this sum corresponds to the maximal support. But after the rule is selected for inclusion in the output set, $c(e)$ values for the examples covered by this rule are increased by 1 (step 8) and in the next iteration rules covering the same examples as the previous rule or rules will be penalized by the smaller sum $\sum 1/c(e)$. In step 9 the selected rule which is included into the output set is eliminated from the generated rule set in order to prevent that the same rule is included more than once into the output set.

For noisy domains the condition that confirmation rules should not cover

Algorithm 1: CONFIRMATION RULE CONSTRUCTION

Input: $E = P \cup N$ (E training set, P target class examples)

Parameters: *min_support* (minimal support),
max_rules (maximal number of generated rules)

Output: set of up to *max_rules* different confirmation rules
for the target class

(1) **generate** (using an association rule learning algorithm [1]) all rules
of the form: $A_{i_1}, A_{i_1}, A_{i_3}, \ldots \rightarrow TargetClass$ with the properties:
a) their confidence is 100%
b) their support in E is higher than or equal to *min_support*

(2) **if** total number of generated rules is equal to or less than *max_rules*
then include all generated rules into the output set and **exit**.

(3) **else**

(4) **for** every $e \in P$ **do** $c(e) \leftarrow 1$

(5) **repeat** *max_rules* times

(6) **select** among generated rules the rule with
the highest sum $\sum 1/c(e)$ where sumation is over the set
$P' \subseteq P$ of the target class examples covered by the rule

(7) **add** the selected rule into the output comfirmation rule set

(8) **for** every $e \in P'$ of the selected rule **do** $c(e) \leftarrow c(e) + 1$

(9) **eliminate** the selected rule from the generated rule set

(10) **end repeat**

(11) **end else**

(12) **exit** with *max_rules* number of constructed rules

Figure 2: Construction of the set of confirmation rules for a selected target class by association search.

any of the non-target class examples may be too strong and this requirement may be relaxed so that the confidence level of accepted association rules can be less than 100%. This presents a simple and practical noise handling approach but it may lead to an uncontrolled reduction of prediction accuracy of induced confirmation rules. Therefore, in experiments in Section 4 we used a procedure for explicit noise detection and elimination in preprocessing [7] which is based on the consensus of saturation filters. The characteristic of this approach is that only a small number of examples with high probability of actually being noisy are detected and eliminated from the training set. This is important for the confirmation rules concept which should provide for a high reliability of decisive predictions.

4 Application of confirmation rules in coronary artery disease diagnosis

Coronary artery disease (CAD) is a disease where one, two or all three coronary arteries are narrowed or obstructed mainly by artherosclerotic plaque(s). The consequence is diminished blood supply causing diminished oxygen supply of the dependent region of the myocardium, manifesting as angina pectoris (AP). The most extreme consequences are myocardial infarction and (cardiac) death.

The dataset, collected at the University Medical Center, Ljubljana, Slovenia, includes 327 patients (250 men and 77 women, mean age 55 years). Each patient had performed history, clinical and laboratory examinations including ECG at rest, ECG during controlled exercise, stress myocardial perfusion scintigraphy, and coronary angiography which gives the diagnosis of coronary artery disease. In 229 patients CAD was angiographycally confirmed and in 98 it was excluded. The patients' clinical and laboratory data are described by 77 attributes. This dataset was previously used for inducing diagnostic rules by a number of machine learning algorithms [8].

The coronary artery disease dataset was used to generate confirmation rules in a series of experiments using different disjoint attribute subsets: symptoms and signs including ECG at rest, ECG during exercise, and myocardial perfusion scintigraphy. These rules may be interesting for disease prediction at various stages of the diagnostic process. In [6] we presented the confirmation rules induced from the complete attribute set. The domain expert evaluated more than half of them as sensible and reliable predictors. Independently, in [7] we used the same domain in order to test the consensus saturation filter on a real medical problem. The results were good because the system detected in total 15 noisy cases (out of 327 patients) out of which the medical doctor who collected the data, recognized 14 as being real outliers, either being errors or possibly noisy cases with coronary angiography tests very close to the border line between the two classes.

In this work we induced confirmation rules using the association rule learning algorithm, adapted to the problem of confirmation rule learning (Algorithm 1). We tested prediction quality of the induced confirmation rules using the 10-fold cross-validation procedure.

In accordance with the standard 10-fold cross-validation procedure, the original data set was partitioned into 10 folds with 32 or 33 examples each. Training sets are built from 9 folds, leaving one fold as a test set. In this way, 10 training sets and 10 corresponding test sets were constructed. Every example occurs exactly once in a test set, and 9 times in training sets. Algorithm 1 was used to construct the confirmation rule sets. In the experiments the *min_support* parameter was set relatively low while *max_rules* parameter was always 5. In this way, for every training set, 5 confirmation rules were generated for the class *not-confirmed* and 5 for the class *confirmed*. Such an experimental setting enabled us to test generated confirmation sets with different acceptance leves. Acceptance level 1 means that it is enough that one of the confirmation rules 'fires' in order to classify the test example. With acceptance levels 2 and 3 it is necessary that at least 2, respectively 3, confirmation rules of the target class are correct for the classification. The prediction is correct (successful) if the example is classified into a single class, which has to be the same as the expert classification. The prediction is erroneous if the example is classified into a single class which is different from the expert classification.

Measured prediction results are presented in Table 1. The table has two parts: the first presents results obtained *without* and the second *with* noise elimination in preprocessing. In both cases results for three different acceptance levels are reported. The first column of every row is the acceptance level followed by the number and percentage of correct predictions and the number and percentage of erroneous predictions. From these numbers the relative measured error rate is calculated. It is known that ischaemic heart disease is a noisy domain, and because of that the measured error rate includes both errors due to the imperfectness of the confirmation rule approach, and 'expected' erroneous predictions due to the domain noise. In order to estimate real error rate (last column in Table 1) we have eliminated from the measured errors 14 examples which were detected and evaluated in [7] as noisy.

Measured error rates are between 3.7% and 9.5% while estimated real error rates are about 0.6% – 4.2% what are the best results for the domain compared with the results obtained both by other machine learning algorithms and medical experts [8, 9]. It must be noted that the elimination of the 'expected' domain noise was extremely conservative, based on the consensus of the saturation filter preprocessor and the domain expert, potentially resulting in overestimation of the real error rate. The least estimated real error rate is detected with acceptance level 3 and noise detection in preprocessing. In this case the number of indecisive predictions is about 50% with only one really wrong prediction in about 150 decisive classifications. This result proves high reliability of the induced confirmation rules, both for the rules generated in this work by the association approach and those already presented in [6].

Results in Table 1 demonstrate also differences among prediction qual-

accept. level	correct predict. total	correct predict. relative	err. predict. total	err. predict. relative	measured rel. err. rate	real rel. err. rate
a) without noise elimination in preprocessing						
1	237	72.48%	25	7.65%	9.54%	4.2%
2	156	47.71%	9	2.75%	5.45%	1.8%
3	93	28.44%	4	1.22%	4.12%	1.0%
b) with noise elimination in preprocessing						
1	249	76.15%	19	5.81%	7.09%	3.2%
2	199	60.86%	10	3.06%	4.78%	2.0%
3	155	47.40%	6	1.83%	3.73%	0.6%

Table 1: Results of 10-fold cross-validation presenting the number of correct predictions, number of erroneous predictions (including total numbers and relative values), measured relative error rate and real relative error rate for **a)** without and **b)** with noise elimination in preprocessing. For each fold with about 294 training cases and 33 test cases, 5 **confirmation rules** for the class *confirmed* and 5 **confirmation rules** for the class *non confirmed* were generated. Results are presented for acceptance levels 1–3, where level 3 means that the case must satisfy at least 3 out of 5 rules for decisive prediction. The relative number of correct predictions represents the total number of correct predicitions divided by the total number of cases (327), while the relative number of erroneous predictions is the total number of erroneous predictions divided by the total number of cases. The measured relative error rate is equal to the ratio of the number of erroneous predictions and the number of all decisive predictions. The real relative error rate is computed so that the number of erroneous predictions is, at first, reduced so that it does not include expert-evaluated domain outliers, and then it is divided by the number of all decisive predictions.

ity for various acceptance levels. As expected, the increased acceptance level reduces the number of correct predictions but it also significantly reduces the number of erroneous predictions, especially real prediction errors. The observation holds with and without noise elimination in preprocessing. Noise elimination itself is very useful. The comparison of the number of correct predictions for confirmation rules generated without and with noise detection and elimination in preprocessing demonstrates the importance of the use of the noise handling mechanism for effective confirmation rule induction. For example, for acceptance level 3 the increase is from 28% to 47%.

5 Conclusion

This work stresses the importance of reliable decision making which is not based on the optimization of prediction accuracy as most frequently used in medical inductive learning applications. For this purpose the paper elaborates the concept of confirmation rules. The proposed framework of confirmation rule induction is general because it enables the incorporation of results of different machine learning algorithms, as well as the existing expert knowledge. The induced structure of an unordered list of simple rules and the possibility of providing predictions of increased reliability are its main advantages. The main disadvantage of the approach are indecisive answers. In presented experiments the number of indecisive predictions has been high, always greater than 20% with a maximum greater than 70%.

In our experiments, association rule learning has proved to be a useful mechanism for constructing confirmation rules. An important characteristic of Algorithm 1 is the induction of confirmation rules with different covering properties. The approach ensures significant diversity of the selected rules although the real independence can not be guaranteed. The conclusion follows from improvements in prediction reliability when higher acceptance levels have been tested.

Acknowledgments

We are grateful to Ciril Grošelj from the University Medical Center, Ljubljana, Slovenia for providing the coronary artery disease dataset and for his involvement in previous experiments in this domain. We are also grateful to Matjaž Kukar from the Faculty of Computer and Information Sciences, University of Ljubljana, for his involvement in the evaluation of results of our previous studies in this domain.

References

[1] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen and A.I. Verkamo (1996) Fast discovery of association rules. In U.M. Fayyad, G. Piatetski-

Shapiro, P. Smyth and R. Uthurusamy (Eds.) *Advances in Knowledge Discovery and Data Mining*, pp. 307–328. AAAI Press.

[2] K.M. Ali and M.J. Pazzani (1996) Error reduction through learning multiple descriptions. *Machine Learning*, 24:173–206.

[3] L. Breiman (1996) Bagging predictors. *Machine Learning* 24(2): 123–140.

[4] Y. Freund and R.E. Shapire (1996) Experiments with a new boosting algorithm. In *Proc. Thirteenth International Machine Learning Conference ICML'96*, 148–156, Morgan Kaufmann.

[5] D. Gamberger (1996) Induction of redundant rules for medical applications. In *Proc. First Workshop on Intelligent Data Analysis in Medicine and Pharmacology IDAMAP-96* at the ECAI-96 conference, 27–30.

[6] D. Gamberger, N. Lavrač, C. Grošelj (1999) Diagnostic rules of increased reliability for critical medical applications. In *Proc. Joint European Conference on Artificial Intelligence in Medicine and Medical Decision Making AIMDM-99*, 361–365, Springer Lecture Notes in AI 1620.

[7] D. Gamberger, N. Lavrač, C. Grošelj (1999) Experiments with noise filtering in a medical domain. In *Proc. of International Conference on Machine Learning ICML-99*, in press.

[8] C. Grošelj, M. Kukar, J.J. Fetich and I. Kononenko (1997) Machine learning improves the accuracy of coronary artery disease diagnostic methods. *Computers in Cardiology*, 24:57–60.

[9] M. Kukar, C. Grošelj, I. Kononenko and J. Fetich (1997) An application of machine learning in the diagnosis of ischaemic heart disease. In *Proc. Tenth IEEE Symposium on Computer-Based Medical Systems*, 70–75, IEEE Computer Society.

[10] R.S. Michalski, I. Mozetič, J. Hong, and N. Lavrač (1986) The multi-purpose incremental learning system AQ15 and its testing application on three medical domains. In *Proc. Fifth National Conference on Artificial Intelligence*, pp. 1041–1045, Morgan Kaufmann.

[11] J.R. Quinlan (1996) Boosting, bagging, and C4.5 . In *Proc. Thirteenth National Conference on Artificial Intelligence*, 725–730, AAAI Press.

[12] R.L. Rivest and R. Sloan (1988) Learning complicated concepts reliably and usefully. In *Proc. Workshop on Computational Learning Theory*, 69–79, Morgan Kaufman.

[13] K.M. Ting and B.T. Low (1997) Model combination in the multiple-data-batches scenario. In *Proc. Ninth European Conference on Machine Learning*, 250–265, Springer Lecture Notes in AI 1226.