Inconsistency tests for patient records in a coronary heart disease database

Dragan Gamberger [1], **Nada Lavrač** [2], **Goran Krstačić** [3], and **Tomislav Šmuc** [4]

Abstract. The work presents the results of inconsistency detection experiments on the data records of an atherosclerotic coronary heart disease database collected in the regular medical practice. Medical expert evaluation of some preliminary inductive learning results have demonstrated that explicit detection of outliers can be useful for maintaining the data quality of medical records and that it might be a key for the improvement of medical decisions and their reliability in the regular medical practice. With the intention of on-line detection of possible data inconsistences, sets of confirmation rules have been developed for the database and their test results are reported in this work.

1 Introduction

The motivation for the research presented in this work stems from the fact that modern medical decision processes are generally based on patient data from many different sources which are typically collected and archived by a multiterminal or distributed computer systems. Such organization enables prompt and high quality decisions of medical doctors supported by abundance of available data [2] but it also enables that errors of different sources, caused by the work of many different people and/or instrumentation can directly enter patient records. Detection of data inconsistencies at the global level of every patient record can help in tracing systematic as well as spurious errors in the data acquisition process and in this way it can be an important part for insuring high quality and reliability of data used for medical decision making [5]. On the other side, the existence of data inconsistencies do not need to be the sign of data errors but may be the consequence of some atypical medical case. Attracting attention of medical doctors to such patient records may be interesting both from the point of view of medical science as well as for avoiding everyday medical practice routine errors.

Our interest in inconsistency testing is the result of many inductive learning experiments performed on a database of atherosclerotic coronary heart disease (ACHD) patients prepared at the Institute for Prevention of Cardiovascular Disease and Rehabilitation, Zagreb Croatia. In the data preparation phase, the saturation filter [4] was used to detect and eliminate outliers from the database. This is a necessary step in the knowledge discovery process which enables the induction of globally relevant rules. Medical expert evaluation has demonstrated that the detection of outliers is a very interesting result by itself, which suggested the idea of using noise detection algorithms developed for data preprocessing in inductive machine learning as a tool for data cleaning of patient records. In this work two different approaches to the problem of inconsistency testing are presented in Section 2. This is followed by the presentation of the medical domain used in experiments and the results of the experiments in Sections 3 and 4, respectively. The algorithms used for outlier detection and rule construction are out of the scope of this work since their description can be found in [3, 4].

2 Inconsistency tests

Machine learning approaches to inconsistency testing in patient records can be either based on outlier (noise) detection algorithms or on a set of rules that are supposed to be true for the data in patient records. The later approach can be used also for on-line inconsistency testing but it requires the construction of rules with specific properties. In both cases, testing is based on supervised machine learning algorithms which require that patient records are grouped in two or more classes. The classes can be defined either by domain experts or by values of one or more descriptors available in the patient record. Correlation between defined classes and data contained in patient records is the main mechanism used in inconsistency detection. Appropriate class assignment is one of the main problems of machine learning approaches to inconsistency testing and it is specially analyzed in Section 4.

2.1 Explicit outlier detection

The first approach, in the work called *explicit outlier detection* can be without changes used on very different patient records. It is actually a noise detection algorithm for data of two classes, described in detail in [4], used in the data preprocessing phase (data cleaning) of inductive learning algorithms. It works on the set of records, trying to identify significant differences among positively and negatively classified records. Complexity of the least complex hypothesis correct for all available examples (true for all positive and false for all negative examples) is estimated without constructing any concrete hypothesis. Those records which are difficult for correct classification and which, by their elimination enable direct reduction of the complexity of the least complex hypothesis, are detected as outliers. The approach is appropriate for off-line data analysis. Its main drawback is its time complexity as well as the fact that for multiclass problems the algorithm must be repeated for every reasonable definition of class positive and class negative records.

[1] Rudjer Bošković Institute, Bijenička 54, 10000 Zagreb, Croatia, E-mail Dragan.Gamberger@irb.hr
[2] Jožef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia, Nada.Lavrac@ijs.si
[3] Institute for Prevention of Cardiovascular Disease and Rehabilitation, Draškovićeva 13, 10000 Zagreb, Croatia
[4] Rudjer Bošković Institute, Bijenička 54, 10000 Zagreb, Croatia, E-mail Tomislav.Smuc@irb.hr

Descriptor	Abbreviation	Characteristics
Anamnestic data		
sex	SEX	1-man 2-woman
age	AGE	continuous (years)
height	H	continuous (m)
weight	W	continuous (kg)
body mass index	BMI	continuous ($kg\ m^{-2}$)
family anamnesis	F.A.	1-negative 2-positive
present smoking	P.S.	1-negative 2-positive 3-very positive
diabetes mellitus	D.M.	1-negative 2-pos. medicament therapy 3-pos. insulin therapy
hypertension	HYP	1-negative 2-positive 3-very positive
stress	STR	1-negative 2-positive 3-very positive
Laboratory tests		
total cholesterol	T.CH.	continuous ($mmol\ L^{-1}$)
trygliceride	TR	continuous ($mmol\ L^{-1}$)
high density lipoprotein	HDL/CH	continuous ($mmol\ L^{-1}$)
low density lipoprotein	LDL/CH	continuous ($mmol\ L^{-1}$)
uric acid	U.A.	continuous ($\mu mol\ L^{-1}$)
fibrinogen	FIB	continuous ($g\ L^{-1}$)

Table 1. The names and the characteristics of 16 anamnestic and laboratory testing descriptors.

2.2 Rule-based outlier detection

The second approach, called *rule-based outlier detection* is more appropriate for on-line inconsistency testing. It works with data of one patient record only and the consequence is its simplicity and high execution speed. The approach is actually a set of logical tests that must be satisfied by every patient record. If one or more of the tests is not satisfied, the record is detected as an outlier. The logical tests are defined by the set of rules that hold for the patient records in the domain. The main drawback of the approach is that the set of rules must be developed specially for the tested type of records. Moreover, the used rules must be highly reliable rules with a very small number of mispredictions, leading to false outlier alarms. Such rules can be constructed by domain experts but also by inductive learning algorithms. Because of the required rule reliability, the concept of *confirmation rules* seems appropriate for this task [3]. In this concept, separate rules are constructed for the positive and negative class cases. The confirmation rules for the positive class must be true for many positive cases and for no negative case. If a negative case is detected true for any confirmation rule developed for the positive class, it is a reliable sign that the case is an outlier. In the same way, confirmation rules constructed for the negative class can be used for outlier detection of positive patient records. An additional advantage of the approach is that the user can have the information about the rule which caused the alarm what can be useful in the error detection process.

3 Data Set

For this work, a database representing typical medical practice in atherosclerotic coronary heart disease ACHD diagnosis has been prepared. The data describe patients who entered the Institute for Prevention Cardiovascular Disease and Rehabilitation, Zagreb Croatia, in a few months period. The set of descriptors represents all potentially interesting and typically available information about patients. The descriptor set includes anamnestic data (10 items), laboratory test results (6), the resting ECG data (5), the exercise test data (5), echocardiogram results (2), vectorcardiogram results (2), and long term continuous ECG recording data (3). It makes altogether 33 descriptors. Only patients with complete data have been included into the data set, resulting in the data set with 238 patients in total. The descriptors are cited in Tables 1 and 2.

Descriptor	Abbreviation	Characteristics
ECG at rest		
heart rate	HR	continuous (beats min^{-1})
ST segment depression	ECGst	1-negative 2-positive 1mm 3-positive $\geq$ 2mm
serious arrhythmias	ECGrhyt	1-negative 2-positive
conduction disorders	ECGcd	1-negative 2-positive
left ventricular hypertrophy	ECGhlv	1-negative 2-positive
Exercise ECG		
ST segment depression	ExECGst	continuous (mm)
serious arrhythmias	ExECGrhyt	1-negative 2-positive
conduction disorders	ExECGcd	1-negative 2-positive
hypertensive reaction	ExECGhyp	1-negative 2-positive
New York Heart Ass. functional class	ExECGNYHA	class I - IV
Echocardiography		
left ventr.		
internal diameter	EchoLVID$_d$	continuous (mm)
left ventr. ejection fraction according to Simpson	EchoLVEF	continuous (%)
Vectorcardiography Q		
transmural MI	VCG Q	1-negative 2-positive
left ventricular hypertrophy	VCGhlv	1-negative 2-positive
Long term continuous ECG		
serious arrhythmias	HOLrhyt	1-negative 2-positive
conduction disorders	HOLcd	1-negative 2-positive
ST segment depression	HOLst	continuous (mm)

Table 2. The names and the characteristics of 17 non-invasive diagnostic descriptors.

The classification of all patients was performed by the cardiologist and it reflects generally accepted medical knowledge. The classification is mostly based on the results of the most important tests. These are: exercise testing, long term ECG recording and echocardiography. In exercise testing ST segment depression or elevation, serious cardiac arrhythmias, and conducting disturbances are the important parameters. Additionally, the NYHA classification [6] and clinically significant metabolic equivalents (METs) are used [1]. Similar parameters can be found in long term ECG recording, except MET and NYHA classification. Dyastolic internal diameter of left ventricular with parasternal short axis view and left ventricular ejection fraction (according to Simpson) are determined from the echocardiogram.

For this research the cardiologist classified patients into 5 groups whose main features are summarized below:

Group I Healthy patients without verified ACHD but with possible present cardiovascular risk factors.

Group II-V These are patients with previous myocardial infarction. They were classified by the results of non-invasive cardiovascular tests and their condition after some coronary angioplastic or cardiosurgery treatment. They are all under medicament treatment.

Group II Patients with normal results of exercise testing, long term recording and echocardiogram.

Group III Patients with ST segment depression 1.00 mm in exercise testing and during long term ECG recording, left ventricular ejection fraction higher than 55%, METs 10.

Group IV Patients with ST segment depression equal or higher than 2.00 mm in exercise testing and during long term ECG recording, left ventricular ejection fraction less than 55% (40-54%), left ventricular internal diameter more than 6.0 cm, NYHA II-III, METs 5–10.

Group V Patients having ST segment elevation or depression $>$ 3.00mm, left ventricular ejection fraction less or equal to 30%, left ventricular inner diameter greater than 6.5 cm, NYHA III-IV, METs$<$5.

In experiments with expert defined classes the groups III-V represented the positive class while groups I-II were in the negative class.

4 Experimental results and medical evaluation

The domain of 238 patient records consists of two sets: the dataset of 150 patients collected earlier, has been used for preliminary experiments and rule development, while the set of remaining 88 records collected later, has been used for test purposes only. The first set will be in the rest of the paper called the main set and the second one the test set.

4.1 Explicit outlier detection results

The set of experiments started with explicit outlier detection for the main set and for the positive and negative classes as defined in Section 3 by medical doctors. There have been only two detected outliers (patient records number 28 and 52) and both of them are very interesting cases. The first one is actually an older patient after a serious cardiosurgical treatment who was in spite of non-optimal laboratory tests intentionally put into Group II (patients with normal results of exercise testing, long term recording and echocardiogram). The second patient was also in Group II but after its detection as an outlier, medical doctor agreed that Group III would be much more appropriate for the patient. In the analysis it was detected that the main reason for its inclusion in Group II were good exercise testing results. But the results were misleading because the patient was so weak that he could sustain only 2 minutes (instead of 7 - 9 minutes) of exercise. It means that an actually important outlier had been detected, that its detection helped in finding more appropriate diagnostic group for the patient, and that the medical doctor has found out that exercise testing results are reliable data only if the tests could be and have been performed completely and correctly.

The same explicit approach to outlier detection was also applied on the test set which resulted in the detection of patient records 174, 214, 227, and 230. More reliable results should be expected if the outlier detection algorithm is applied on the target dataset consisting of both main and test set. The result of this experiment was the detection of the same four detected records from the test set and in total five records from the main set. Besides examples 28 and 52, that were detected also in the first experiment, the set of detected records from the main set included also cases number 1, 43, and 98. Table 3 summarizes the results of the first three experiments with explicit outlier detection. The results show a weakness of the explicit approach to noise detection demonstrated by the fact that different outliers have been detected for the main set depending on the target set.

Target set	**Detected outliers**	
	In main set	**In test set**
main	28 52	- - -
test	- - -	174 214 227 230
main + test	1 28 43 52 98	174 214 227 230

Table 3. Results of explicit outlier detection for different sets of target patient records.

Medical evaluation showed that cases 1 and 43, patients from Group III, are not expert recognized outliers. These cases can be accepted as false alarms of the explicit outlier detection approach. A completely different situation is the patient number 98 from Group III, classified as a serious case but with practically normal results of laboratory tests. A medical doctor accepted the patient as a special case and was satisfied that machine learning methods recognized it as an outlier. Due to the patient's medical history, the case classification remained unchanged. Two out of four cases detected in the test set are border line cases (cases number 174 and 227) and the remaining two are real medical outliers: one is a difficult coronary patient from Group V with diagnosed cardiopathia dilatativa therefore, not an ACHD patient (number 214), while the other one is an atypical ACHD patient whose disease could be detected only by echocardiography (number 230).

4.2 Rule-based outlier detection results

With the intention to show the application of a rule-based outlier detection approach, the main set was used to induce confirmation rules, explained in Section 2.2, for both classes. The rules for the positive class were $ExECGst > 0.45mm$ and $HOLst > 0.65mm$. Each of these two rules is true for about 95% of the positive class cases in the main set and false (except for explicitly detected outliers) for all negative class cases in this set. These two rules can be used as constraints that should not be satisfied by negative class cases. The rules detected the following negative class outliers in the test set: 165, 174, 185, and 227. There was only one confirmation rule for the negative class consisting of two conditions $ExECGst \leq 0.45mm \wedge HOLst \leq 0.65mm$. The rule is true for about 98% of negative class cases and false for all positive class cases in the main set. The rule detected positive patient records 214 and 230 as outliers within the test set. The results are presented in the first row of Table 4. Comparing results obtained by explicit and rule-based outlier detection it can be noted that the later approach selected the same records as the former approach but that it also detected two more records: 165 and 185. The result demonstrates the applicability of the rule-based method for inconsistency testing. The method is interesting because its application is much simpler for all other future patient records. Medical evaluation of the cases 165 and 185 showed that they are actually not false alarms. One of them is a border line case who was intentionally put in Group II because of its medical history (case number 185) while in the other case the medical doctor accepted the suggestion and he has changed the patient group classification (case 165).

Classes defined by	Outlier detection Explicit	Rule-based
medical doctors	174 214 227 230	165 174 185 214 227 230
$ExECGst > 0.45mm$	165 174 185 195 197 199 202 227 232 237	165 174 185 195 197 199 202 227 232 237
$HOLst > 0.65mm$	165 185 195 197 199 202 214 230 232 237	165 185 195 197 232

Table 4. Comparison of the outliers detected by explicit and rule-based detection approaches for differently defined patient record classes.

4.3 Results obtained by descriptor-based classifiers

In all previous experiments inconsistency testing was based on classes defined by medical doctors. Although the results are very reasonable for both methods and they agree in most detected records, this way of inconsistency testing is appropriate only for domains in which doctor classification exists. Existence of such classification assumes that there exists a dependency between the determined class and record data what practically ensures the quality of the inconsistency tests. In a general case, when there is no expert classification, one or more descriptors from the patient record should be used as a classification parameter. In this situation it is essential to select classifiers for which it can be assumed that their dependency with other data in the record exists. In the ACHD domain the induced rules demonstrated strong dependencies between expert classes and ST segment depression during exercise and long term continuous ECG monitoring. This is the main reason for selecting these data as appropriate classifiers when expert classification does not exist. The limit values between positive and negative classes are based on induced rules in previous experiments. Table 4 in its second and third row includes results obtained for positive classes defined by conditions $ExECGst > 0.45mm$ and $HOLst > 0.65mm$ respectively. In the left column are records detected as outliers in the test set by explicit approach while in the right one are detected by the rule-based method. In the explicit approach, the target sets included both the main and the test sets. Rules for the rule-based detection have been constructed always from the main set so that its own outliers have been previously excluded from it. The method resulted in successful detection of the most outliers detected based on expert classification. It should be noted that among outliers occur also some completely new cases, like numbers 195, 197, 232, and 237. Their medical evaluation demonstrated that in three out of four cases the problem was that the patient exercise testing was incomplete and the obtained results were misleading. In the fourth case (number 195) the patient had an asymptomatic (silent) ischaemic heart disease known by its differences between exercise and long term continuous measurements. It must be noted that detection of these four outliers was medically completely justified. Analysis of medical classifications in all four cases showed that medical expert reasoning also successfully detected the problems and that all cases were in appropriate groups in spite of data inconsistencies.

4.4 Subconcept discovery using outlier detection

The rules induced in the previous experiments show that both approaches for outlier detection use dependencies of a small number of data from the record. This practically means that only inconsistencies in a relative small part of the record can be detected. The problem can be solved by using the suggested methods iteratively based on same patient classifiers but with different descriptor subsets. The results of experiments in the ACHD domain with different descriptor subsets were not consistent and their reasonable medical evaluation was difficult. The cause of the problem can be the insufficient dependency among less important data in the records.

Some experiments with descriptor subsets led to very interesting results, typically detecting small patient subsets with special properties. In one of them all potential outliers in the positive class were characterized by positive reaction to drug therapy (i.e. their risk factor parameters were inside normal limits) as compared to a majority of ill patients that did not react positively to the drug therapy. Detection of this subset is an example of subconcept discovery which might be interesting for medical research purposes. In this way, our approach could be used in the same way as a subgroup discovery system.

5 Conclusions

This paper introduces two approaches to outlier (inconsistency) detection in medical datasets: explicit outlier detection and rule-based outlier detection. Explicit outlier detection is applicable in off-line analysis since it operates on the dataset level and involves data cleaning algorithms usually used in machine learning preprocessing. Rule-based outlier detection relies on rules induced upon previously collected data in the same dataset. It is applicable for on-line detection of inconsistencies in future records. We have applied both approaches on the ACHD patient dataset. Experiments were performed with expert classified records as well as with different descriptor-based classifications. The results indicated the sensitivity of both approaches for inconsistency detection. Although detected outliers differed from one experiment to another, most outliers were confirmed as special cases by subsequent medical expert evaluation. In order to detect possible inconsistencies in less important descriptors, experiments with different descriptor subgroups have been performed. Medical evaluation in some of these experiments recognized outliers with similar characteristics representing interesting domain subconcepts. The results could be important both with respect to everyday medical practice as well as for future medical research.

REFERENCES

[1] ACP/ACC/AHA Task force on Exercise Testing (1990). *Journal of American College Cardiology* **16**: 1061-1065.

[2] Diamond, G.A.,& Forester, J.S. (1979). Analysis of probability as an aid in the clinical diagnosis of coronary artery disease. *New England Journal of Medicine* **300**:1350.

[3] Gamberger, D., Lavrač, N., and Grošelj C. (1999) Diagnostic rules of increased reliablity for critical medical applications. In *Proc. of Joint European Conference on Artificial Intelligence in Medicine and Medical Decision Making (AIMDM'99)*, pp.361–365.

[4] Gamberger, D., Lavrač, N., and Grošelj C. (1999) Experiments with noise filtering in a medical domain. In *Proc. of International Conference of Machine Learning (ICML'99)*, pp. 143–151.

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

[6] Wayne A.R., Schlant R.C., Fuster V., In HURST'S: The Heart, Arteries and Veins. McGrawc Hill, NY, 1127-1263.