

MULTIMETHOD MACHINE LEARNING APPROACH FOR MEDICAL DIAGNOSING *

M. Lenič¹, P. Povalej¹, M. Zorman¹, V. Podgorelec¹, P. Kokol¹, L. Lhotska²

¹ Laboratory for system design, Faculty of Electrical Engineering and Computer Science, University of Maribor, Smetanova 17, Maribor, Slovenia
mitja.lenic@uni-mb.si

² Gerstner Lab, Czech Technical University in Prague, Faculty of Electrical Engineering, Czech Republic
lhotska@feld.cvut.cz

Abstract- In this paper we present a multimethod approach for induction of a specific class of classifiers, which can assist physicians in medical diagnosing in the case of mitral valve prolapse. Mitral valve prolapse is one of the most controversial prevalent cardiac condition and may affect up to ten percent of the population and in the worst case results in sudden death. MultiVeDec is a general framework enabling researchers to generate various intelligent tools based on machine learning. In this paper we focused on various decision tree methods, which are capable of extracting knowledge in a form closer to human perception, a feature that is very important in medical field. The experiment included classifiers with various classical single method approaches, evolutionary approaches, hybrid approaches and also our newest multimethod approach. The main concern of the latest approach is to find a way to enable dynamic combination of methodologies to the somehow quasi unified knowledge representation. The proposed multimethod approach was capable to outperform all other tested approaches by producing classifier for diagnosing mitral valve prolapse with the highest overall and average class accuracy.

Keywords - intelligent system, decision making, diagnosing, mitral valve prolapse, multimethod approach.

I. INTRODUCTION

Many real-world medical problems are nowadays being treated with tools for automatic intelligent data analysis. Various methods have been developed to improve the quality of analysis for specific domains. Application of any method in a specific domain requires special characteristics, like for example instance methods based on artificial neural networks are capable of generalisation of nonlinearly separable problems, but have poor explanatory power not suitable for human understanding. While medical experts are not "very good accustomed with numbers" we as informaticians focused on methods, which are capable of extracting knowledge in a form closer to human perception, e.g. methods that induce decision trees, classification rules, etc.

In the project presented in this paper we focus on inducing an user friendly classifier, which can assist physicians in diagnosing mitral valve prolapse (MVP) [1]. That is very important while MVP is one of the most

prevalent cardiac conditions, which may affect from five up to ten percent of adult population, and is also one of the most controversial ones, and thereafter hard to diagnose without the use of expensive technology normally not available to general practitioners.

With a selection of knowledge representation and concentrating on decision trees we have only narrowed down the potential set of methods. Thereafter in order to find most appropriate classifier from above we tried out different methods for decision tree induction. First we made analysis with widely known tools for decision tree induction C4.5 and C5/See5 [2]. Despite acceptable results we wanted to find alternative solutions i.e. evolutionary and multimethod approach.

Main contribution of our paper are:

- the throughout analysis and comparison of different decision tree induction methods on a medical case
- the introduction of the multimethod decision tree induction method
- the introduction of the general multimethod approach.

II. MITRAL VALVE PROLAPSE

Prolapse is defined as the displacement of a bodily part from its normal position. The term mitral valve prolapse (MVP)[1][3][4], therefore, implies that the mitral leaflets are displaced relative to some structure, generally taken to be the mitral annulus. The silent prolapse is the prolapse which can not be heard with the auscultation diagnosis and is especially hard to diagnose. The implications of the MVP are the following: disturbed normal laminar blood flow, turbulence of the blood flow, injury of the chordae tendinae, the possibility of thrombus' composition, bacterial endocarditis and finally hemodynamic changes defined as mitral insufficiency and mitral regurgitation.

Mitral valve prolapse is one of the most prevalent cardiac conditions, which may affect from five up to ten percent of population and is one of the most controversial one. The most common cause is probably myxomatous change in the connective tissue of the valvular leaflets that makes them excessively pliable and allows them to prolapse into the left atrium during ventricular systole. The clinical manifestations

* The research in this paper was supported by the grant of Slovenian Ministry of Science, Education and Sport: Slovenian - Czech cooperation - Project ISCoMAA

of the syndrome are multiple. The great majority of patients are asymptomatic. Other patients, however may present atypical chest-pain or supraventricular tachyarrhythmias. Rarely, patients develop significant mitral regurgitation and, as with any valvular lesions, bacterial endocarditis is a risk.

Uncertainty persists about how it should be diagnosed and about its clinical importance. Historically, MVP was first recognized by auscultation of a mid systolic "click" and late systolic murmur, and its presence is still usually suggested by auscultatory findings. However, the recognition of the variability of the auscultatory findings and of the high level of skill needed to perform such an examination has prompted a search for reliable laboratory methods of diagnosis. M-mode echocardiography and 2D echocardiography have played an important part in the diagnosis of mitral valve prolapse because of the comprehensive information they provide about the structure and function of the mitral valve.

Medical experts propose [1][3][4] that echocardiography enables properly trained experts armed with proper criteria to evaluate mitral valve prolapse (MVP) almost 100%. Unfortunately however, there are some problems concerned with the use of echocardiography. The first problem is that current MVP evaluation criteria are not strict enough. The second problem is the incidence of the MVP in the general population and the unavailability of the expensive ECHO - machines to general practitioners. According to above problems we have decided to develop a decision support system enabling the general practitioner to evaluate the MVP using conventional methods and to identify potential patients from the general population.

III. METHODS

Machine learning community has a long tradition in knowledge extraction that can be traced at least as far as the mid-1960. Through the time different approaches evolved [5], such as symbolic approaches, computational learning theory, neural networks, etc. Most of the strength was and is concentrated in finding a way to extract generalized knowledge from the examples.

The selection of appropriate method for analysis of data can be crucial for success as we have shown in following sections. Therefore, for a given problem, different methods should be applied to increase quality of extracted knowledge. Brief overview of selected methods is presented in the following subsections.

A. Classical approaches

Decision trees [2] are easy understandable to the human and can be used even without a computer, but they have difficulties expressing complex nonlinear problem. On the other hand, connectivistic approaches, that simulate cognitive abilities of the brain, can extract complex relation, but are not understandable to humans, and therefore in such way not directly usable for data mining.

There are many other approaches, like representation of the knowledge with rules, rough-sets, case based reasoning,

support vector machines, different fuzzy methodologies, ensemble methods [6] and they all try to answer the question: How to find optimal solution, i.e. learn how to learn.

B. Evolutionary approach

Evolutionary approaches to knowledge extraction are also a good alternative, because they are not inherently limited to local solution. They are based on the evolutionary ideas of natural selection and genetic processes of biological organisms. As the natural populations evolve according to the principles of natural selection and "survival of the fittest", first laid down by Charles Darwin, so by simulating this process, genetic algorithms are able to evolve solutions to real-world problems, if they have been suitably encoded [7]. They are often capable of finding optimal solutions even in the most complex of search spaces or at least they offer significant benefits over other search and optimisation techniques.

C. Hybrid approach

The hybrid approaches rest on the assumption that only in the synergetic combination of single models can unleash their full power [8]. Each of the single method has its advantages, but also inherent limitations and disadvantages, which must be taken into account when using the particular method. Therefore the logical step is to combine different methods to overcome the disadvantages and limitations of a single method.

D. Multimethod approach

While studying presented approaches we were inspired by the idea of hybrid approaches and evolutionary algorithms. Both approaches are very promising in achieving the goal to improve the quality of knowledge extraction and are not inherently limited to sub-optimal solutions. We also noticed that almost all attempts to combine different methods use loose coupling approach. The methods work almost independent of each other and therefore a lot of luck is needed to make them work as a team.

Each of those methods uses its own internal knowledge representation (symbolic, connectivistic) that other methods cannot reuse, because of the incompatibility of knowledge representations. That incompatibility presents a major obstacle when trying to combine different methods using conventional hybrids.

But as already mentioned above, the idea of knowledge exchange is not new. There has been a lot of research going on in the extraction of knowledge from neural nets and vice versa. Although conversions are not ideal, the majority of knowledge can be transformed from one form to another giving us the possibility to exchange knowledge between different methods. With proper combination of those different methods we expect to significantly improve

extracted knowledge. Opposed to the conventional hybrids described in the previous section, our idea is to dynamically combine and apply different methods in not predefined order to the same problem or the decomposition of the problem.

Another aspect of knowledge exchange is that methods have to be able to accept already constructed knowledge representation, and have to apply its operations with the objective to improve the quality of the already extracted knowledge, or they have to be able to construct a knowledge representation from the scratch.

Main concern of the multimethod approach [9] is to find a way to enable dynamic combination of methodologies to the somehow quasi unified knowledge representation. Multiple equally qualitative solutions like in EA approach, where each solution is gained using application of different methodologies with different parameters were used. Therefore we introduced a population composed out of individuals/solutions that have the common goal to improve their classification abilities on a given environment/problem. We have also enabled coexistence of symbolic and cognitive representation in the same population. The most common knowledge representation models have to be standardized to support the use different methods on individuals. In that manner the transformation support between each individual method does not need to be provided. The action is based on the assumption that it is highly improbable to find unified representation for all knowledge representations, therefore we decided to standardize the most popular representations like neural nets, decision trees, rules, etc. Standardization brings in general greater modularity and interchangeability, but it has following disadvantages - already existing methods cannot be directly integrated and have to be adjusted to the standardized representation.

Usually methods are composed out of operations that can be reused in other methods. Therefore we introduced the operation on an individual, a function that transforms one or more individuals to a single individual. Operation can be a part of one or more methods, like pruning operator, boosting operator, etc. Operator based view provides us with the ability to simply add new operations to the framework.

The representation with individual operations facilitate an effective and modular way to represent the result as single individual, but in general the result of operation can be a population of individuals (for example mutation operation in EA is defined on individual level and on the population level). Therefore population operations that generally accept a population as the input and return the population as the result were introduced. The single method itself is composed out of population operations that use individual operations and is introduced as a strategy in the framework that improves individuals in a population. Population operators can be generalized with higher order function and thereby reused in different methods.

To increase the modularity and extensibility of the framework the idea of object oriented paradigm has been used. The polymorphism and inheritance in operations and individual representations has been introduced.

TABLE I
MVP CLASSIFICATION ACCURACY (\square), AVERAGE CLASS ACCURACY (Δ) ON TEST SET AND RANK OF THE METHOD BASED ON AVERAGE CLASS ACCURACY

Method	\square	Δ	Rank
Greedy linear	91.54	83.47	2
Greedy Boost linear	91.54	74.51	16
Nested Boost	93.08	62.54	36
Genetic	92.31	65.83	32
Multimethod	93.84	84.31	1
Greedy Voting	90.00	78.43	10
Greedy ID3	88.46	73.39	19
Boost Greedy ID3	91.54	70.94	20
Greedy Chi square	90.00	82.91	7
Boost Greedy Chi square	90.77	57.21	41
Greedy Gini	88.46	77.87	14
Boost Greedy Gini	87.69	64.15	35
Greedy J measure	83.08	50.84	53
Boost Greedy J measure	89.23	53.08	49
Greedy Mgain	87.69	68.63	28
Boost Greedy MGain	85.39	55.25	46
Greedy Chi square + ID3	91.54	83.47	2
Boost Greedy Chi square + ID3	89.23	56.65	42
Greedy Chi square * ID3	90.77	70.66	21
Boost Greedy Chi square * ID3	93.08	70.59	22
Greedy Gini + ID3	92.31	79.27	8
Boost Greedy Gini + ID3	90.77	69.75	25
Greedy Gini * ID3	90.00	78.43	10
Boost Greedy Gini * ID3	92.31	57.77	40
Greedy Gini + Chi square	91.54	83.47	2
Boost Greedy Gini + Chi square	89.23	56.65	42
Greedy Gini * Chi square	87.69	69.54	27
Boost Greedy Gini * Chi square	90.00	61.42	37
Greedy J measure + ID3	86.15	51.96	50
Boost Greedy J measure + ID3	89.23	48.60	56
Greedy J measure * ID3	81.54	50.28	54
Boost Greedy J measure * ID3	90.77	53.64	48
Greedy J measure + Chi square	88.46	74.30	17
Boost Greedy J measure + Chi square	91.54	66.46	30
Greedy J measure * Chi square	83.85	55.60	45
Boost Greedy J measure * Chi square	87.69	48.95	55
Greedy J measure + Gini	85.39	51.68	51
Boost Greedy J measure + Gini	89.23	40.55	59
Greedy J measure * Gini	80.77	41.95	57
Boost Greedy J measure * Gini	89.23	40.55	59
Greedy MGain + ID3	88.46	77.87	14
Boost Greedy MGain + ID3	90.77	69.75	25
Greedy MGain * ID3	90.77	78.71	9
Boost Greedy MGain * ID3	92.31	70.31	23
Greedy MGain + Chi square	90.77	83.19	6
Boost Greedy MGain + Chi square	85.39	59.73	39
Greedy MGain * Chi square	89.23	78.15	12
Boost Greedy MGain * Chi square	89.23	56.65	42
Greedy MGain + Gini	87.69	68.63	28
Boost Greedy MGain + Gini	85.39	55.25	46
Greedy MGain * Gini	89.23	78.15	12
Boost Greedy MGain * Gini	90.00	61.42	37
Greedy MGain + J measure	80.00	41.67	58
Boost Greedy MGain + J measure	90.77	66.18	31
Greedy MGain * J measure	73.08	51.68	51
Boost Greedy MGain * J measure	87.69	31.93	61
C4.5	91.50	65.50	33
C5	91.50	65.50	33
Boost C5	92.30	70.00	24
IBR - all attributes	77.69	73.95	18
IBR - reduced(10) attributes	91.54	83.47	2

We extended the idea with aspect oriented paradigm, that enables clear separation of concerns and avoids tangled individual representation. With this approach we achieved the modularity and extensibility of the framework, which does not impose to many constrains to the implementation of methods. The individual and population operations can be easily waved together with no additional effort.

IV. RESULTS AND DISCUSSION

Using the Monte Carlo sampling method, 900 children and adolescents were selected representing the whole population under eighteen years of life. Routinely they were called for an echocardiography no matter of prior findings. 631 of them passed an examination of their health state in a form of a carefully prepared protocol specially made for the syndrome of MVP. The protocol consisted of 103 parameters that can possibly indicate the presence of MVP. Distribution of the three decision classes were: 5% "prolapse", 6% "silent prolapse", and 89% "no prolapse".

The following basic purity metrics for greedy induction of decision trees where used: Information gain ratio (ID3), χ^2 (Chi square), Gini, J Measure [11,12,13], IBR (iBARET) [14] and newly developed Mgain metric. iBARET represents a learning tool based on principles of instance-based learning and is derivative of the well-known kNN method. iBARET implements a batch feature weighting method with performance bias, i.e., it uses feedback from the performance function during training. This function is calculated for each setting of feature weights and a genetic algorithm is used to suggest new sets of feature weights. Receiver operating curve methodology is used to calculate this fitness function, the methodology is able to optimize classifier performance in domains with non-uniform class distribution.

Different linear combinations of classic purity metrics were introduced and evaluated. In addition AdaBoost method [13] was used in order to improve the quality of classically induced classifiers. In order to make objective assessment of classifier quality we used average class accuracy, which on unbalanced data sets is more informative measure than overall accuracy. The classification results on testing set are presented in table I.

The methods in table I were ranked according to the average class accuracy of the induced classifier. It can be seen, that the average class accuracy varied from 31.93% (Boost Greedy MGain * J measure) in the worse case to 84.31% in the best case (multimethod). On the other hand the overall accuracy varied from 73.08% (Greedy MGain * J measure) to 93.84% (multimethod). Considering both evaluation criteria we established that the average class accuracy is far more distinctive compared to overall accuracy. Nevertheless, the multimethod approach outperformed other methods in both categories. Explanation for better performance of multimethod approach can be found in the concept of multimethod approach. Classical approaches always produce one deterministic result that is not necessarily an optimal solution. Evolutionary approach overcomes limitation of classical approaches and can find

global optima, but solution search space is huge and the optima is not necessarily found. Multimethod approach tries to combine benefits of classical and evolutionary approaches and reduces search space with help of conventional methods.

V. CONCLUSION

In the paper we presented methods for induction of classifiers that can assist physicians in diagnosing mitral valve prolapse. Mitral valve prolapse is one of the most controversial prevalent cardiac condition and may affect up to ten percent of population. The aim of the research was to identify important factors for diagnosis. The multimethod approach outperformed other methods on mitral valve prolapse database.

REFERENCES

- [1] R. Devereux, "Diagnosis and Prognosis of Mitral Valve Prolaps", *The New England Journal of Medicine*, vol. 320, no. 16 1989, pp. 1077-1079.
- [2] J.R. Quinlan. *C4.5: Programs for Machine Learning*, Morgan Kaufmann publishers, San Mateo, 1993.
- [3] H.R. Anderson et al, "Clinicians Illustrated Dictionary of Cardiology", *Science Press*, London, 1991.
- [4] J.B. Barlow et al, "The Significance of Late Systolic Murmurs", *American Health Journal*, vol. 66, pp. 443-452, 1963.
- [5] S. Thrun and L. Pratt, *Learning to Learn*. Kluwer Academic Publishers, 1998
- [6] T.G. Dietterich, "Ensemble Methods in Machine Learning". in *First International Workshop on Multiple Classifier Systems, Lecture Notes in Computer Science* New York: Springer Verlag, 2000, pp. 1-15.
- [7] D.E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison Wesley, Reading MA, 1989.
- [8] C.J. Iglesias, "The Role of Hybrid Systems in Intelligent Data Management: The Case of Fuzzy/neural Hybrids", *Control Engineering Practice*, vol. 4, no. 6, 1996, pp. 839-845.
- [9] M. Lenič, P. Kokol, "Combining Classifiers with Multimethod Approach", in *Hybrid intelligent systems*, in press.
- [10] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, *Classification and Regression Trees*, Wadsworth International Group, Belmont, CA, 1984.
- [11] G.W. Snedecor and W.G. Cochran, *Statistical Methods*, Eighth Edition, Iowa State University Press, 1989.
- [12] M.A. Bramer, "Using J-pruning to reduce overfitting in classification", in *Knowledge Based Systems*, vol. 15, issues 5-6, pp. 301-308, July 2002.
- [13] Y. Freund, R.E. Schapire, "Experiments with a new boosting algorithm", in *Proceedings Thirteenth International Conference on Machine Learning*, Morgan Kaufman, San Francisco, pp. 148-156, 1996.
- [14] J. Klema, J. Palouš, "iBARET - Instance-Based Reasoning Tool", *Eunite Proceedings*, Verlag Mainz, pp. 213-221, 2001