Intelligent Analysis of Long-Term Patient Survival after Pacemaker Implantation

P Povalej[1], M Krizmaric[2], P Kokol[1], Z Pehnec[3]

[1]University of Maribor, Faculty of Electrical Engineering and Computer Science, Smetanova ulica 17, 2000 Maribor, Slovenia
[2]University of Maribor, College of Nursing Studies, Zitna ulica 15, 2000 Maribor, Slovenia
[3]General Hospital of Maribor, Department of Emergency Medicine, Ljubljanska 2, SI-2000 Maribor, Slovenia

Abstract

In the recent years intelligent systems proved to be a useful and successful tool which has often been used for decision support, data mining and knowledge discovery in medicine. Data mining and knowledge discovery enable to prove existing hypothesis or to generate new ones – thus they in longer term support the advancement of medicine as a scientific discipline and as a consequence developing new treatment, prediction and diagnosing methods directly helping clinicians.

In this paper we present the results of intelligent data analysis used for determining important factors for long-term patient's survival after pacemaker implantation. The study included 2207 patients which had their first pacemaker implantation during 1966 and 2001 in Slovenia.

1. Introduction

Cardiac pacemakers have been in clinical use more than 30 years. Early devices provided only single-chamber, asynchronous and nonprogrammable pacing. Single-chamber pacemakers functioned at a fixed rate, generally about 70 beats per minute and regardless of the patients activity level and blood pumping requirements. Today, advanced electronics afford dual-chamber multi programmability, diagnostic functions, rate response, data collection, reliability, and lithium-iodine power sources extend longevity to upward of 10 years [1]. This types of pacemakers contained a tiny sensor that detected pressure waves caused by a patient's muscle movement or body motion; the pacemaker's circuitry translated those waves into electrical signals, which in turn triggered the appropriate rate response from the pacemaker. Continual advances in a number of clinical, scientific, and engineering disciplines have so expanded the use of pacing that it now provides cost-effective benefits.

The idea of the study presented in this paper is to examine the important factors that can influence the long-term patient's survival after pacemaker implantation. We will focus on intelligent data analysis tools which have already been frequently used to prove existing hypothesis or to generate new ones – thus they in longer term support the advancement of medicine as a scientific discipline and as a consequence developing new treatment, prediction and diagnosing methods directly helping clinicians.

Concerning above, intelligent decision support tools are developed, which are able to process a huge amount of data available from solving previous cases and suggesting the probable diagnosis or extracting knowledge based on the values of several important attributes. Clearly, black-box classification methods (neural networks for example) are not appropriate for this kind of task, because the clinical experts need to evaluate and validate the decision making process, induced by those tools, before there is enough trust to use the tools in practice. On the other hand, the evaluation of the induced classifiers produced by the computerized tools by a clinical expert can be an important source of new knowledge on how to make a diagnosis based on the avail-able attributes. In order to achieve this goal, the classification process should be easily understandable and straightforward. Different kinds of knowledge discovery methods are therefore appropriate to do the job but we decided to use only the ones that are based on the decision trees since they provide a very important feature – the possibility of explaining the decisions in a way understandable by humans. Therefore we developed a new multimethod approach for decision tree induction, which in general combines different classic approaches with genetic algorithms and is more precisely presented in Section 2.

In this paper we present the results of using multimethod approach for knowledge extraction on real-world database of patients with pacemaker implantation in Slovenia. The database is presented in Section 3 and in Section 4 we present the results. The paper concludes with some general remarks and directions for further research.

2. Method

Machine learning community has a long tradition in knowledge extraction that can be traced at least as far as the mid-1960. Trough the time different approaches evolved [2], 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 will show in the following sections. Therefore we developed a new multimethod approach, which attempts to successfully combine classical and genetic approaches. Brief overview of methods used is presented in the following subsections.

2.1. Classical approach

Decision trees [3] 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 method-ologies, ensemble methods [4] and they all try to answer the question: How to find optimal solution, i.e. learn how to learn.

2.2. 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 algo-rithms are able to evolve solutions to real-world problems, if they have been suitably encoded [4]. 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.

2.3. Hybrid approach

Hybrid approaches rest on the assumption that only in the synergetic combination of single models can unleash their full power [6]. Each of the single method has its ad-vantages, 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.

2.4. Multimethod approach

While studying presented approaches we were inspired by the idea of hybrid ap-proaches and evolutionary algorithms. Both approaches are very promising in achiev-ing 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 independ-ent 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 mutlimethod approach [7] is to find a way to enable dynamic combination of methodologies to the somehow quasi unified knowledge representa-tion. 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 environ-ment/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 trans-forms 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 facilitates 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.

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.

Multimethod approach combines advantages of single methods and avoids theirs disadvantages at the same time by applying different methods on the same knowledge model, each of which may contain inherent limitations, with the expectation that the combined multiple methods may produce better results. The result is presented in population of knowledge representations for the specified problem that also enables different point of view (second opinion) to given problem.

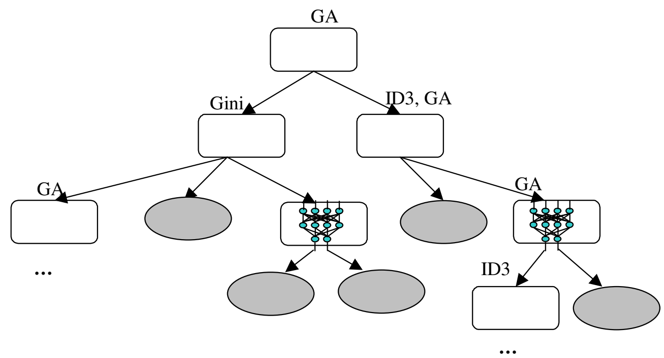

Figure 1- An example of a decision tree induced using multimethod approach. Each node is induced with appropriate method (GA – genetic algorithm, ID3, Gini, Chi-square, J-measure, SVM, neural network, etc.)

3. The study

The study included 2207 patients which had their first pacemaker implantation during 1966 and 2001 in Slovenia. The patients records contained the data about: the year of birth, gender, the year of the first implantation, the age at the first implantation, number of implantations, mode of the pacemaker, indulgence and survival in years. The data analysis showed that almost 50% of patients lived from 2-5 years after the implantation and 26% of patients died in the first year. There was no essential difference in survival length according to gender. The importance of the patient's age at the time of the first implantation was revealed. Almost 50% of the patients had their pacemaker in VVI pacing mode, 22.3% in VVIR, 16.46% in DDD, 6.16% in VDD, 5.84% in DDDR and the rest 0.96% of the patients in AAIR pacing mode.

4. Results

In order to perform the intelligent data analysis a complete dataset was randomly divided into training and testing set at a ratio 2 to 1.

In the first experiment we divided the survival of the patients in four classes: less than one year, one to five years, five to ten years and more than ten years. We tested the data on several different methods but the accuracy of classification didn't go higher than 69%. However, due to so many output classes those kinds of results were expected. Therefore the number of classes was diminished to three with combining the last two classes.

Consequently an interesting fact was discovered – the most important attribute in induced decision trees was year of implantation and the “key” year was 1992. Most of the patients with pacemaker implantation before the year 1992 lived less than 5 years.

This results lead to new experiment in which we filtered the database according to the patient’s age at first implantation - we excluded all patients that were older than 79 years. We also filtered out all patients that had their first implantation after 1992 and simultaneously had the last check in the hospital after 2001 (because they were alive at the end of this study). After the filtering the new database contained 515 patients out of which 23 lived less than one year and 325 lived more than five years. Thereafter we decided to join the first two classes into one. The new distribution of output classes was the following: 36.9% of patients that survived less than five years and 63.1% of patients with more or equal to five years survival rate. The most interesting decision tree was 89.5% accurate on training data and 80% accurate on test data. Again the year 1992 was most frequently used as the first attribute in the decision tree. Most of the patients, which had their first pacemaker implantation in 1992 or after, lived at least five years. For the patients operated before 1992 the next important attribute in the decision tree was the number of implanted pacemakers - patients with more than one pacemaker implantation also lived longer.

In order to obtain better distribution between output classes in the last experiment we decided to change the boundary line to less or equal to five years (43.3%) and more than five years of long-time survival (56.7%). As expected some other interesting attributes were shown in induced multimethod decision trees. The most powerful attribute was number of implanted pacemakers. The majority of patients with more than one pacemaker implantation lived more than five years. However, when the patients with only one pacemaker implantation were considered the next important feature was the mode of pacemaker. The decision tree with 81% accuracy on the learning set and 78% accuracy on the testing set revealed that pacemakers with rate modulation enable longer (more than five years) survival.

5. Discussion and conclusions

We presented the results of intelligent data analysis used for determining the important factors in a long-term patient survival after permanent pacemaker implantation. The study included 2207 patients which had their first pacemaker implantation during 1966 and 2001 in Slovenia. Intelligent data analysis using advanced multimethod approach was performed for classification of the long term survival. For that purpose a filtered database was used in order to obtain more accurate results.

All decision trees were evaluated on the basis of accuracy and average class accuracy of classifying unseen test cases.

Some interesting new patterns were revealed during different experiments. One of the most frequently shown patterns was that patients operated in 1992 had higher survival rate. Number of pacemaker implantations was also one of the most powerful attributes. A highly accurate decision tree revealed that patients with only one pacemaker implantation had higher survival rate if the pacemaker was programmable (NBG IV Code = R - Rate Modulation).

However, the decision trees with the highest predicting power have to be more precisely examined also by clinical experts so that the results can be evaluated in practice.

References

[1] Bronzino JD. Biomedical Engineering Handbook, Boca Raton. CRC Press, Heidelberg : Springer, [Piscataway] : IEEE Press; 2000.

[2] Thrun S, Pratt L. Learning to Learn. Kluwer Academic Publishers 1998.

[3] Quinlan JR. C4.5: Programs for Machine Learning. San Mateo: Morgan Kaufmann publishers, 1993.

[4] Dietterich TG. Ensemble Methods in Machine Learning. In: First International Workshop on Multiple Classifier Systems, Lecture Notes in Computer Science. New York: Springer-Verlag, 2000: 1-15.

[5] Goldberg DE. Genetic Algorithms in Search, Optimization, and Machine Learning. Addison Wesley Reading MA 1989.

[6] Iglesias CJ. The Role of Hybrid Systems in Intelligent Data Management: The Case of Fuzzy/neural Hybrids, Control Engineering Practice 1996; 4 (6): 839-845.

[7] Lenič M, Kokol P. Combining classifiers with multimethod approach. In: Ajith et al. ABRAHAM. Soft computing systems: design, management and applications 2002, Frontiers in artificial intelligence and applications. Amsterdam : IOS Press , 2002; 87: 374-383.

Address for correspondence

Petra Povalej
Faculty of Electrical Engineering and Computer Science, University of Maribor
Smetanova ulica 17, 2000 Maribor
Slovenia
Petra.Povalej@uni-mb.si