Journal of Information Technologies, Vol. 7, No. 1, 2014, ISSN: 1337-7467

Use of a propositional rule learner for prognosis of mortality rates in heart failure patients

Jan Bohacik[1], Department of Computer Science at the University of Hull, UK and Department of Informatics at the University of Žilina, Slovakia

C. Kambhampati, Department of Computer Science at the University of Hull, UK

Darryl N. Davis, Department of Computer Science at the University of Hull, UK

John G. F. Cleland, Department of Cardiology at the University of Hull, UK

Abstract: Nowadays, as a consequence of increasing costs and rising numbers of heart failure patients, hospitals are trying to optimise their capacity by telemonitoring patients in an outpatient setting. An important part of telemonitoring applied to heart failure patients is a clinician decision support system that is able to prognosticate at any moment if a particular patient could die within a certain period of time. Use of a propositional rule learner using data about patients with heart failure is investigated for this type of prognosis with the aim of minimizing life-threatening situations while maintaining the costs of treatment. Achieved experimental results regarding the performance of the rule learner are provided.

Key words: propositional rule learner, prediction, heart failure, telemonitoring

1. Introduction

Telemonitoring transmits parameters such as weight or pulse rate for review by health-care professionals. It is promoted as one of possible and low-cost solutions for the management of patients with heart failure worldwide. Although results of its use are contradictory, meta-analyses have suggested that telemonitoring may provide better clinical outcomes than usual care, with a reduction in mortality and hospital admissions observed (13). Importance of telemonitoring is likely to increase since the number of patients with heart failure is rising quickly due to factors such as aging population, obesity and diabetes (11) and this is accompanied by escalating health care costs.

---

[1] Ján Boháčik with all Slovak diacritics.

An important future is prognosis of a patient's possibility to die within some time once telemonitoring is established as a solution for treatment of patients with heart failure. This future is meant for the purposes of active prevention for telemonitored patients. An issue here is a lack of methods that could do this type of prognosis. Some clinical methods such as EFFECT Risk Scoring system (9), Emergency Heart Failure Mortality Risk Grade (EHMRG) (10) or Seattle Heart Failure Model (SHFM) (8) are considerable for this type of task. Since telemonitoring can potentially provide enormous amounts of collectable data about patients with heart failure, data mining and machine learning appear to be very useful methods. There are several examples of data mining and machine learning applied in medical domains (1)(2)(3)(12)(14). In this article, a propositional rule learner based on (7) is used for prognosis of mortality rates in heart failure patients.

The article has the following organization. The patient data used in the study is described in Section 2. In Section 3, our application of a propositional rule learner for prognosis of mortality rates in heart failure patients is presented. Section 4 provides the achieved experimental results. Section 5 contains conclusions of the study.

2. Used patient data

The patient data used in the study contains information about heart failure patients $\boldsymbol{p} \in \boldsymbol{P}$ where $|\boldsymbol{P}| = 2032$ and it is a part of Hull LifeLab - large, epidemiologically representative, information-rich clinical data (6). The patients are described by nine attributes $\boldsymbol{A}$ **=** **{**A_1**; ...;** A_k**; ...;** A_9**}** and each patient $\boldsymbol{p} \in \boldsymbol{P}$ has one assigned prognosis $c_j \in C = \{c_1; c_2\}$. More details can be found in Table. 1. If A_k is a discrete attribute, $A_k = \{a_{k,1}; ...; a_{k,l}; ...; a_{k,l_k}\}$ where $a_{k,1}$, ...; $a_{k,l}$, ..., a_{k,l_k} are possible discrete values. *Weight* (A_1) is the patient's weight in kilograms. *Height* (A_2) is the patient's height in meters. *Sex* (A_3) shows if the patient is female or male. *Age* (A_4) is the patient's age in years. *Pulse Rate* (A_5) is the rate of the patient's pulse measured by tactile on the outside of an artery in beats per minute. *NT-proBNP Level* (A_6) represents the amount of the N-terminal prohormone of brain natriuretic peptide (NT-proBNP) in picograms per millilitre of the patient's blood. *Blood Creatinine Level* (A_7) is the amount of creatinine in micromoles per litre of the patient's blood. *Blood Sodium Level* (A_8) is the amount of sodium in millimoles per litre of the patient's blood. *Blood Creatinine Level* (A_9) is the amount of creatinine in micromoles per litre of the patient's blood. Value *alive* (c_1)

means that the patient is prognosticated to be alive at least more than six months. Value *dead* (c_2) is prognosticated to be dead within six months.

Table. 1: Patient data.

Attribute	**Type**	**Obtained values**
Weight (A_1)	Continuous	29.8000 – 193.8000 (kg)
Height (A_2)	Continuous	1.2000 – 1.9600 (m)
Sex (A_3)	Discrete	*female* ($a_{3,1}$)
		male ($a_{3,2}$)
Age (A_4)	Continuous	27.0000 – 96.0000 (years)
Pulse Rate (A_5)	Continuous	38.0000 – 150.0000 (bpm)
NT-proBNP Level (A_6)	Continuous	0.8900 – 18236.0000 (pmol/L)
Blood Uric Acid Level (A_7)	Continuous	0.1100 – 1.0600 (mmol/L)
Blood Sodium Level (A_8)	Continuous	123.0000 – 148.0000 (mmol/L)
Blood Creatinine Level (A_9)	Continuous	37.0000 – 1262.0000 (umol/L)
Prognosis (C)	Discrete	*alive* (c_1)
		dead (c_2)

3. Propositional rule learner for prognosis of mortality rates

The prognosis is based on a set of rules R used together with current values of attributes in $\boldsymbol{A}$ for a patient $\boldsymbol{p}$, i.e. all $A_k(\boldsymbol{p})$, $A_k \in \boldsymbol{A}$. R is first created with a particular algorithm and then it is supposed to be used for prognosis for a longer period of time. The following algorithm (so-called propositional rule learner) is used for creation of R (7)(4):

> Set $R = \emptyset$ and for both $c_j \in C$ from the less prevalent one to the more frequent one, execute:
>
> > 1. Building stage:

Repeat 1.1 and 1.2 until the descrition length of R and patients $\boldsymbol{p} \in \boldsymbol{P}$ is 64 bits greater than the smallest descrition length met so far, or there are no positive patients, or the error rate >= 50%.

1.1. Grow phase:

Grow one rule by greedily adding antecedents (or conditions) to the rule until the rule is as accurate as possible. Every possible value of each attribute is tried and the condition with highest information.

1.2. Prune phase:

Incrementally prune each rule and allow the pruning of any final sequences of the antecedents; The pruning metric is (p-n)/(p+n) where p is the number of c_2 and n is the number of c_1. However, it is actually 2p/(p+n) - 1, so in the implementation the following is used: (p+1)/(p+n+2). Thus if p+n is 0, it is 0.5.

2. Optimization stage:

After generating the initial rules $R_i \in R$, generate and prune two variants of each rule R_i from randomized data using procedure 1.1 and 1.2. But one variant is generated from an empty rule while the other is generated by greedily adding antecedents to the original rule. Moreover, the pruning metric used here is (tp+tn)/(p+n) where tp is the number of true positives and tn is the number of true negatives. Then the smallest possible descrition length for each variant and the original rule is computed. The variant with the minimal descrition length is selected as the final representative of R_i in R. After all $R_i \in R$ have been examined and if there are still residual positives, more rules are generated based on the residual positives using Building Stage again.

3. Delete the rules from R that would increase the descrition length of the whole R if it were in it. Distributions of values in C are also computed for each rule in R.

Prognostication for a particualr patient $\boldsymbol{p}$ is as follows (5):

For each rule r in R execute:

If r covers values $A_k(\boldsymbol{p})$, $A_k \in \boldsymbol{A}$, value $c_j \in C$ with the maximal value in the distribution related to rule r is chosen as the result. STOP.

4. Experiments

The purpose of the experiments was to show the performance of the analysed propositional rule learner on the patient data. The experiments were conducted with a Java software tool developed by the first author. The core algorithm of the propositional rule learner was implemented in Weka (15) as class JRip. The performance was measured with sensitivity, specificity, positive predictive value, negative predictive value, and accuracy defined as tp / (tp + fn), tn / (tn + fp), tp / (tp + fp), tn / (tn + fn), and (tp + tn) / (tp + fp + fn + tn), respectively. Tp/tn/fp/fn is the number of true positives/true negatives/false positives/false negatives where "*C* is *alive*" is considered negative and "*C* is *dead*" is considered positive. Tp, fp, fn and tn are computed in 10-fold cross-validation where the patient data is partitioned into 10 folds of patients. Of the 10 folds, a single fold is retained as the testing data for evaluation of performance, and the remaining 9 folds are used as the learning data. The learning data is used for building of a rule set. The cross-validation process is repeated 10 times, with each of the 10 folds used exactly once as the testing data. It should not be prognosticated that soon-to-be death patients remain alive as this leads to possible life-threatening situations in telemonitoring. Similarly, it should not be prognosticated that patients remaining alive are going to die as this leads to increased running costs of telemonitoring. Since the former can be measured by sensitivity and the letter by specificity, the sum of sensitivity and specificity is a measure of minimizing life-threatening situations while maintaining the costs of treatment. The higher the value of the sum, the better the performance is.

Table. 2: Experimental results regarding the propositional rule learner.

Measure	**Value**
Sensitivity	0.3615
Specificity	0.9180
Sensitivity + Specificity	1.2795

Positive Predictive Value	0.6026
Negative Predictive Value	0.8070
Accuracy	0.7756

The results showing the performance are available in Table. 2. For the propositional rule learner used on the patient data explained in Section 2, sensitivity is 0.3615, specificity is 0.9180, and the sum of sensitivity and specificity is 1.2795.

5. Conclusions

A propositional rule learner was used for creation of a rule set using data about patients with occurred heart failure. This rule set allowed us to prognosticate the death of a patient within six months. The data consisted of 2032 patients which were alive or dead six months after the values of nine attributes about each of them were acquired. The performance of the rule set created with the propositional rule learner was evaluated in 10-fold cross-validation where measures sensitivity, specificity, positive predictive value, negative predictive value, and accuracy were computed. The prognosis was conducted with the aim of minimizing life-threatening situations while maintaining the costs of treatment. The life-threatening situations are associated with sensitivity and the costs are associated with specificity. The higher their values are, the better the prognosis is. The achieved sensitivity was 0.3615, the achieved specificity was 0.9180, and the sum of sensitivity and specificity was 1.2795. A higher sensitivity and specificity might be achieved with improvements of the propositional rule learner, for example, uncertainties in heart failure data could possibly be incorporated using notions of fuzzy logic.

6. Bibliography

(1) Bohacik, J., Davis, D. N.: Data mining applied to cardiovascular data, Journal of Information Technologies (Volume: 3, Issue: 2), Pages: 14-21, Year: 2010.

(2) Bohacik, J., Kambhampati, C., Davis, D. N., Cleland, J.: Alternating decision tree applied to risk assessment of heart failure patients, Journal of Information Technologies (Volume: 6, Issue: 2), Pages: 25-33, Year: 2013.

(3) Candelieri A., Conforti D., Perticone F., Sciacqua A., Kawecka-Jaszcz K., Styczkiewicz K.: Early detection of decompensation conditions in heart failure patients by knowledge discovery: the HEARTFAID approaches, Proceedings of Computers in Cardiology, Pages: 893-896, Year: 2008.

(4) Documentation of Weka: Class JRip, Retrieved August 2013, from http://weka.sourceforge.net/doc.dev/weka/classifiers/rules/JRip.html.

(5) Source Code of Weka: weka.classifiers.rules.Jrip, Retrieved August 2013, from http://grepcode.com/file/repo1.maven.org/maven2/nz.ac.waikato.cms.weka/weka-stable/3.6.6/weka/classifiers/rules/JRip.java

(6) Clinical Effectiveness and Evaluation Unit of the Royal College of Physicians: Managing chronic heart failure: learning from best practice (Published by: The Lavenham Press Ltd; In: United Kingdom), Pages: 48, Year: 2005.

(7) Cohen W. W.: Fast Effective Rule Induction. Proceedings of the Twelfth International Conference on Machine Learning, Pages: 115-123, Year: 1995.

(8) Ketchum E. S., Jacobson A. F., Caldwell J. H., Senior R., Cerqueira M. D., Thomas G. S., Agostini D., Narula J., Levy W. c.: Selective improvement in Seattle Heart Failure Model risk stratification using iodine-123 metaiodobenzylguanidine imaging, Journal of Nuclear Cardiology (Volume: 19; Number: 5), Pages: 1007-1016, Year: 2012.

(9) Lee D. S., Austin P. C., Rouleau J. L., Liu P. P., Naimark D., Tu J. V.: Predicting mortality among patients hospitalized for heart failure: derivation and validation of a clinical model, JAMA (Volume: 290; Number: 19), Pages: 2581-2587, Year: 2003.

(10) Lee D. S., Stitt A., Austin P. C., Stukel T. A., Schull M. J., Chong A., Newton G. E., Lee J. S., Tu J. V.: Prediction of heart failure mortality in emergent care: a cohort study, Annals of Internal Medicine (Volume: 156; Number: 11), Pages: 767-775, Year: 2012.

(11) Lopez-Sendon J.: The heart failure epidemic, Medicographia (Volume: 33; Number: 4), Pages: 363-369, Year: 2011.

(12) Phillips K. T., Street W. N.: Predicting outcomes of hospitalization for heart failure using logistic regression and knowledge discovery methods, Proceedings of AMIA Annual Symposium, Pages: 1080, Year: 2005.

(13) Polisena J., Tran K., Cimon K., Hutton B., McGill S., Palmer K., Scott R. E.: Home telemonitoring for congestive heart failure: a systematic review and meta-analysis. Journal of Telemedicine and Telecare (Volume: 16), Pages: 68–76, Year: 2010.

(14) Trnka, A., Kovarova, M., Doci, I.: Occurrence of sexual disorders in Slovakia - a five-year analysis with Data Mining, Forum Staticum Slovacum (Volume: 7; Number: 2), Pages: 185-192, Year: 2011.

(15) Witten I. H., Frank E., Hall M. A.: Practical machine learning tools and techniques (Third Edition) (Published by: Morgan Kaufman; In: United States of America), Pages: 664, Year: 2011.

7. Addresses of the authors:

Eur Ing Dr Jan Bohacik
Department of Computer Science/Department of Informatics
Faculty of Science and Engineering/Faculty of Management Science and Informatics
University of Hull/University of Žilina
HU6 7RX/010 26
Hull/Žilina
United Kingdom/Slovakia
J.Bohacik@hull.ac.uk/Jan.Bohacik@fri.uniza.sk

Dr C. Kambhampati
Department of Computer Science
Faculty of Science and Engineering
University of Hull
HU6 7RX
Hull
United Kingdom
C.Kambhampati@hull.ac.uk

Dr Darryl N. Davis
Department of Computer Science
Faculty of Science and Engineering
University of Hull
HU6 7RX
Hull
United Kingdom
D.N.Davis@hull.ac.uk

Professor John G. F. Cleland
Department of Cardiology
Castle Hill Hospital
University of Hull
HU16 5JQ
Hull
United Kingdom
J.G.Cleland@hull.ac.uk

Acknowledgments

The study was carried out within a HEIF-5 funded project. The data about heart failure patients was provided by the Department of Cardiology, Hull York Medical School, University of Hull, UK as a part of Hull LifeLab.

The article was accepted for publication in February 2014 by the publisher of the Journal of Information Technologies (Vol. 7, No. 1, ISSN: 1337-7467), i.e. by the Faculty of Mass Media Communication, University of Ss. Cyril and Methodius in Trnava, Slovakia.