Knowledge representation methods,
machine learning, medical diagnosis

Bartłomiej ŚNIEŻYŃSKI*

CHOICE OF A KNOWLEDGE REPRESENTATION METHOD FOR LEARNING CLASSIFIERS IN MEDICAL DOMAINS

Knowledge based systems in medical domains are common nowadays. Machine learning techniques are broadly used to generate knowledge for such systems. Developers have to choose not only the learning method, but also, what is even more important, the knowledge representation method. The most common criterion for such a choice is prediction accuracy. In the paper we argue that in certain cases knowledge representation, and its simplicity and intelligibility, are more important. In this paper results of experiments performed using several medical data sets and chosen machine learning algorithms are presented. Next, some examples of learned classifiers are shown. Analysis of results conclude the work.

1. INTRODUCTION

Medicine is traditionally very popular domain for Artificial Intelligence (AI) applications. A lot of research is performed. There are many scientific journals, which main subject is AI application in medicine. The research is not purely theoretical, there are also many working knowledge based systems. Descriptions of several dozens of expert systems in this field can be found at [5].

In such systems knowledge can be represented in two forms: symbolic or non-symbolic. The most popular symbolic methods are decision trees, decision rules, and probabilistic methods, where Bayesian networks are most commonly used. The most popular non-symbolic method is application of artificial neural networks.

It is well known that the biggest problem during preparation of a knowledge base system is a knowledge acquisition. This is why machine learning is one of the main streams in AI nowadays. Hundreds of versions of knowledge generation algorithms were developed so far. Every one has some advantages and disadvantages. For most of them it is possible to find data sets for which it will have better accuracy then other algorithms. But is the accuracy the most important factor when one has to choose a learning algorithm and a knowledge representation for a given task? The thesis of this paper is that in some cases appropriate knowledge representation is more important then accuracy.

In the paper we present experimental results of knowledge generation for publicly available medical data sets using decision tree learning, rule induction, Bayesian networks, and multilayered perceptron. Next, some examples of learned classifiers are presented. Analysis of results conclude the work.

* AGH University of Science and Technology, Computer Science Department, Krakow, Poland

2. DATA SETS

Data sets used in experiments are obtained from Machine Learning Repository maintained at University of California, Irvine [4]. There are about one hundred data sets from various domains in this repository. For the purpose of this paper, the following eight sets from medical domain are selected:

- Audiology – original owner is Professor Jergen at Baylor College of Medicine,
- Wisconsin Brest Cancer – obtained from Dr. William H. Wolberg, University of Wisconsin Hospitals, Madison [7],
- Diabetes – donated by Vincent Sigillito, The Johns Hopkins University, Laurel,
- Heart disease – collected by Andras Janosi, Hungarian Institute of Cardiology, Budapest,
- Hepatitis – donated by G. Gong, Carnegie-Mellon University, Pittsburgh,
- Hypothyroid – supplied by the Garavan Institute and R. Quinlan, New South Wales Institute, Syndney,
- Lympography – provided by M. Zwitter and M. Soklic, University Medical Centre, Institute of Oncology, Ljubljana
- Primary tumor – as above.

All these datasets contain records describing patients. For every patient several attributes and a diagnosis (class) is stored. Attributes can be nominal or numeric. It happens that some attribute values are unknown. Short summaries of these sets are presented in Tab. 1.

Table 1. Data sets

#	Data set	# classes	# attributes	#nominal attr.	#numeric attr.	missing val.	#instances
1	Audiology	24	70	70	0	some	226
2	Breast Cancer (W)	2	9	0	9	little	699
3	Diabetes	2	8	0	8	none	768
4	Heart disease (H)	2	13	7	6	many	294
5	Hepatitis	2	19	13	6	some	155
6	Hypothyroid	4	29	22	7	some	3772
7	Lymphography	4	18	15	3	none	148
8	Primary tumor	22	17	17	0	some	339

3. ALGORITHMS AND SOFTWARE USED

As it was mentioned above, there is a lot of machine learning algorithms that can be used to build a classifier from a data. Because we have chosen specific knowledge representation methods, the following popular algorithms that generate knowledge in these representations are used in experiments:

- J48 – implementation of Quinlan's C4.5 [8], well known decision tree generation algorithm,
- Estimator classes for Naive Bayes – used to generate simplified Bayesian network, where a node representing a class is the only parent for all other nodes representing attributes [6],
- K2 algorithm for learning structure of Bayesian network [2], and simple estimator to generate conditional probabilities of learned structure directly from data,
- Jrip – implementation of Cohen's RIPPER [1], one of the most popular rule induction algorithms,
- Back propagation algorithm – used to train multilayered sigmoid perceptron [9], with one hidden layer consisting of (*number of attributes* + *number of classes*) / 2 units,

– ZeroR – very simple algorithm that creates classifier predicting the most common class [10], which is used to set a base line for accuracy.

All experiments were performed using WEKA Data Mining Program [10] and default settings for classifiers. The only exception was setting maximal number of parents equal 10 instead of 1 for Bayesian network generation. Discretization of numeric attributes was done automatically by a program when it was necessary.

4. EXPERIMENTAL RESULTS

Experiments were performed using data split. Randomly selected 70% of instances was used for training models, remaining 30% of instances was used for testing. Procedure was repeated 10 times.

4.1. ACCURACY

Average accuracy and its standard deviation values are presented in Tab. 2. Absolute values of accuracy can be misleading, because in most cases we get more then 50% correct answers using ZeroR classifier. To show more informative vales, ZeroR accuracy was subtracted from other classifiers values. So defined accuracy gain is presented in Fig. 1. One can notice that for some data sets (e.g. Hypothyroid data set) for which we have very high accuracy, we have low accuracy gain. It means that the knowledge generated does not contribute a lot.

Table 2. Experimental results: average accuracy and its standard deviation (best accuracy is bolded)

#	Dataset	J48		Naive Bayes		Bayes Net		Jrip		Neural Net		ZeroR	
		Acc.	Std.d.	Acc.	Std.d.	Acc.	Std.d.	Acc.	Std.d.	Acc.	Std.d.	Acc.	Std.d.
1	Audiology	79.24	3.68	71.51	4.47	74.25	3.11	72.77	3.29	**81.99**	3.56	24.87	1.19
2	Breast Cancer (W)	94.53	0.68	96.06	1.07	**96.63**	0.56	95.77	1.20	95.91	0.84	65.49	0.20
3	Diabetes	72.13	3.17	75.26	3.13	**76.70**	2.57	75.74	3.01	76.30	2.54	65.10	0.15
4	Heart disease (H)	77.64	4.66	**84.81**	3.44	84.14	2.36	79.87	2.13	77.66	4.96	54.46	0.41
5	Hepatitis	79.77	3.33	82.81	2.34	**83.23**	3.26	79.78	3.80	81.73	4.31	79.58	0.90
6	Hypothyroid	**99.40**	0.14	95.64	0.38	99.36	0.25	99.26	0.28	94.36	0.45	92.28	0.08
7	Lymphography	77.17	6.82	82.38	4.10	**83.94**	4.68	76.30	6.28	82.61	5.39	54.53	0.79
8	Primary tumor	41.89	3.33	**48.34**	3.36	44.31	3.90	38.65	2.19	43.26	4.96	24.86	0.67

What is surprising, in all domains, accuracy of all algorithms was similar. Difference between the best and the worst classifier was not higher then 10%.

Another interesting result is that Naive Bayes, that is a subclass of Bayesian network performs almost as good as Bayesian network, and in two cases even better. See [3] for discussion of this phenomenon and some proposed modifications to learning Bayesian networks that increase accuracy of this classifier.

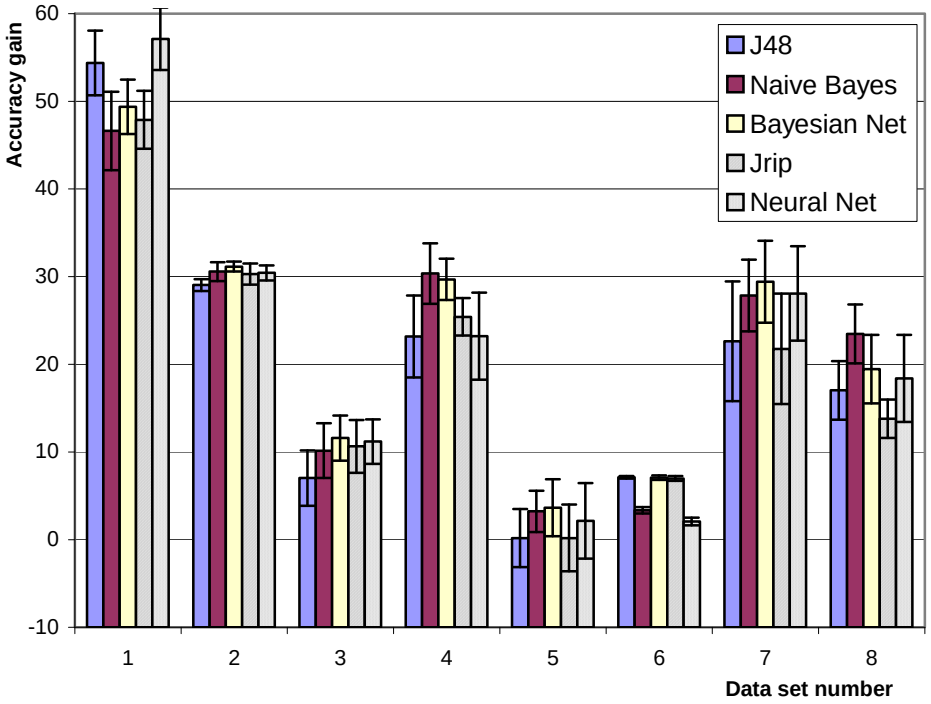

Fig. 1. Accuracy gain

4.2. KNOWLEDGE

In this section some examples of classifiers generated in experiments for Lymphography data set are presented.

Structure of naive Bayes is simple; however, the model consist of about one hundred conditional probabilities. Bayesian network structure is more complicated. Its example is presented in Fig. 2. This model also contains about one hundred real numbers, which are difficult to grasp. Similar problem appears when one wants to analyze neural network, which is also described by similar amount of numbers – connection weights.

Decision tree generated using J48 algorithm is presented in Fig. 3. It is quite complex; however it can be analyzed by a human. Rules generated by Jrip are presented in Fig. 4. As we can see, rule representation is the simplest and the easiest to grasp.

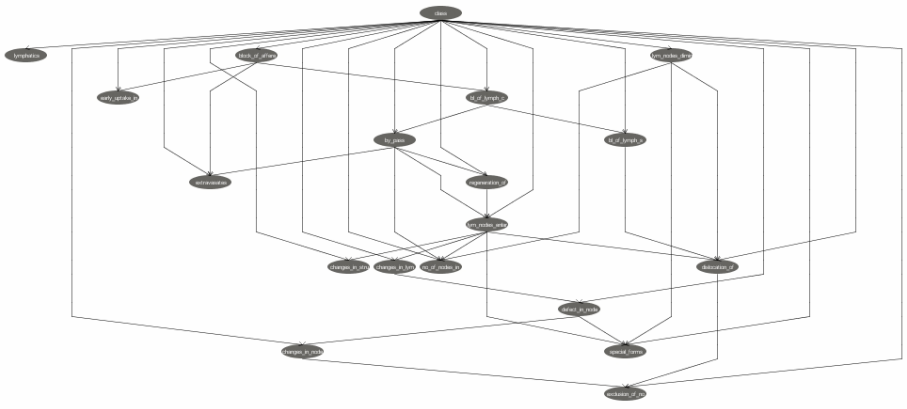

Fig. 2. Example of Bayesian network structure for Lymphography data set

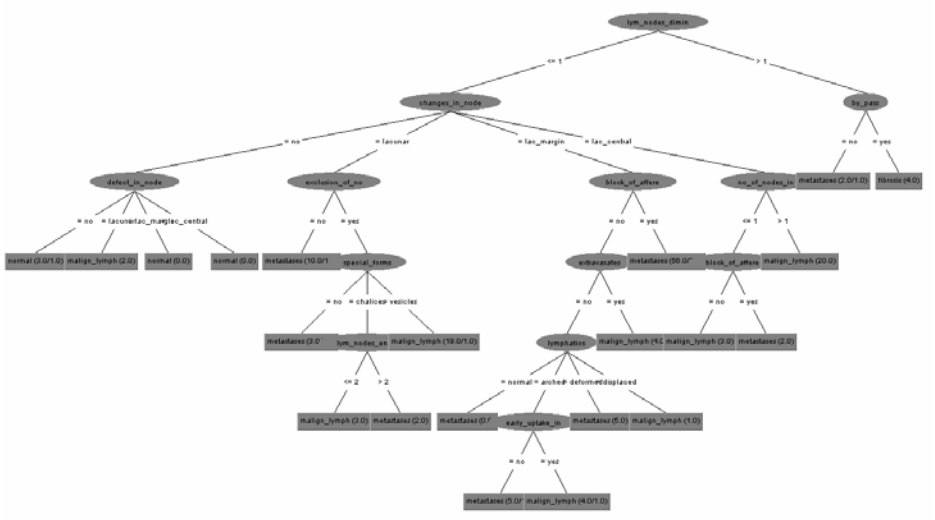

Fig. 3. Example of decision tree for Lymphography data set

```
(lymphatics = normal) → class=normal
(lym_nodes_dimin >= 2) and (by_pass = yes) → class=fibrosis
(no_of_nodes_in >= 3) and (special_forms = vesicles) → class=malign_lymph
(block_of_affere = no) and (extravasates = yes) → class=malign_lymph
(changes_in_node = lac_central) → class=malign_lymph
→ class=metastases
```

Fig. 4. Example of decision rules for Lymphography data set

5. CONCLUSIONS AND FURTHER RESEARCH

Using accuracy factor only, probabilistic methods seem to be the best choice for learning classifiers. Especially, when decision is made by the computer system autonomously. But there are certain situations when accuracy is not the most important.

Often, before generated knowledge is used in an important domain, it should be verified by a human expert. It is hardly possible to verify a knowledge that has form, which is difficult to grasp.

In practise, generated classifiers are often used with a supervision of a human. It is vital especially in a medical domain, where output of the classifier can affect a health or even a life of a patient. To verify the system's decision, the supervisor should have a possibility to check and understand the justification of the answer.

What is also very important, using machine learning one can discover a new knowledge. Nevertheless, it can not be done without feedback from a human expert. To make it possible, the knowledge has to have a form that is easy to interpret for humans.

In these circumstances, decision trees, and decision rules seem to be a better choice then probabilistic and also non-symbolic methods. As we can see, when one has to choose a knowledge representation for a given task, accuracy is not the only factor that should be considered.

Further research will consider building hybrid classifiers that would combine advantages of classifiers with the highest accuracy and ones with natural knowledge representation. Another direction is to develop methods of knowledge transformation that could be used to transform knowledge into various models.

BIBLIOGRAPHY

[1] COHEN, W.W., Fast Effective Rule Induction. Machine Learning: Proceedings of the Twelfth International Conference, A. Prieditis and S. Russell (Edts.), pp. 115–123, Morgan Kaufmann, 1995.
[2] COOPER, G., HERSKOVITZ, A Bayesian method for the induction of probabilistic networks from data. Mach. Learn., Vol. 9, No. 4, pp. 309-347, 1992.
[3] FRIEDMAN, N, GEIGER, D., GOLDSZMIDT, M., Bayesian Network Classifiers, Machine Learning, Vol. 29, pp. 131–163, 1997.
[4] HETTICH, S., BLAKE, C.L., MERZ, C.J., UCI Repository of machine learning databases, http://www.ics.uci.edu/~mlearn/MLRepository.html, University of California, Department of Information and Computer Science, Irvine, CA, 1998.
[5] http://medexpert.imc.akh-wien.ac.at/
[6] JOHN, G.H., LANGLEY, P., Estimating Continuous Distributions in Bayesian Classifiers. Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence. pp. 338-345. Morgan Kaufmann, San Mateo, 1995.
[7] MANGASARIAN, O. L., WOLBERG, W. H., Cancer diagnosis via linear programming, SIAM News, Vol. 23, No. 5, pp. 1-18, 1990.
[8] QUINLAN, R., C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, San Mateo, CA, 1993.
[9] TADEUSIEWICZ, R., Sieci neuronowe, Warszawa 1993.
[10] WITTEN, I.H., FRANK, E., Data Mining: Practical machine learning tools and techniques, Morgan Kaufmann, San Francisco, 2005.