JOURNAL OF MEDICAL INFORMATICS & TECHNOLOGIES Vol. 17/2011, ISSN 1642-6037

Bayes decision theory, multistage classifier, medical decision support systems

Robert BURDUK[1]

SELECTION OF CLASSIFIER IN ACUTE ABDOMINAL PAIN DIAGNOSIS WITH DECISION TREE MODEL

The article presents the application of the decision tree classifier to the acute abdominal pain diagnosis. The recognition task model is based on a decision tree. In this model the decision tree structure is given by the experts. For the assumed structure of the decision tree the locally optimal strategy is considered. The problem discussed in the work shows a selection of different classifiers (parameters) to the internal nodes of the decision tree. Experiments conducted for selected medical diagnosis problem shows that the use of different parameters for k-NN classification can improve the quality of classification in comparison with the situation if it is used with the same parameter for all internal nodes of the decision tree.

1. INTRODUCTION

The concept of a decision support system for diagnosis in acute abdominal pain is not new. Many papers are corresponded to this diagnosis problem. One of the first applications in the field of diagnosis of acute abdominal pain was presented in the work [3]. The rule-based learning systems [6,8] and other approach [1,8] are well suited for Diagnosis in acute abdominal pain. This paper focuses on using different classifiers (parameters) in each node of decision tree problem.

The content of the work is as follows: Section 2 introduces idea of Bayesian approach, idea of top down induction of decision tree, and short introduce into the methods used during computer experiments. In the next section we describe mathematical model of the acute abdominal pain decision problem. Then we present results of the experimental investigations of the proposed decision tree. The last section concludes the paper.

2. THE MULTISTAGE RECOGNITION TASK

The basic idea involved in a multistage approach is to break up a complex decision into several simpler classifications [12]. A decision tree classifier and the hierarchical one are two possible approaches to multistage pattern recognition. Hierarchical classifiers are a special type of multistage classifiers which allows rejection of class labels at intermediate stages. The synthesis of hierarchical classifier is a complex problem. It involves specification of the following components [10]:

1) design of a decision tree structure,
2) selection of features used at each non-terminal node of decision tree,
3) choice of decision rules for performing the classification.

Let us present shortly ideas of the above mentioned two approaches. The first of them uses given decision tree structure and a set of features for each tree's node. This method focuses its attention on decision rule construction based on Bayesian approach for each node. The second group uses different methods of tree induction to return univariate trees or multivariate ones.

3. BAYESIAN HIERARCHICAL CLASSIFIER

Firstly let us shortly present the idea of classifier based on the Bayes rule. Among different concepts and methods of using "uncertain" information in pattern recognition, an attractive from the theoretical point of view and efficient approach is the use of the Bayes decision theory. This approach

[1] Department of Systems and Computer Networks, Faculty of Electronics, Wroclaw University of Technology, Wybrzeze Wyspianskiego 27, 50-370 Wroclaw, Poland.

consists in assumption [5] that the feature vector $x=(x^{(1)},x^{(2)},...,x^{(d)})$ (describing the object being under recognition) and class label $j\in\{1,2...,M\}$ (the object belonged to) are the realization of the pair of the random variables *X*, *J*. In medical application the area *X* describes the result of patient examinations and *J* denotes the patient state. Random variable *J* is described by the prior probability p_j, where

$$p_j = P(J=j) \tag{1}$$

X has probability density function

$$f(X=x \mid J=j) = f_j(x) \tag{2}$$

for each *j* which is named conditional density function. These parameters can be used to enumerating posterior probability according to Bayes formulae:

$$p(j|x)=\frac{p_j f_j(x)}{\sum_{j=1}^{M} p_j f_j(x)} \tag{3}$$

The formalisation of the recognition task leads to the setting of a optimal Bayes decision algorithm $\Psi(x)$, which minimizes expected value of so-called loss function which describes cost of wrong classification [5]. For well known 0-1 loss function mentioned classifier assures the lowest value of the probability of misclassification, and the decision rule chooses the class for which the posterior probability achieved the biggest values [4]

$$\Psi(x)=i \text{ if } p(i|x)=\max_{l\in\{1,...,M\}} p(l|x) \tag{4}$$

In the real situation the prior probabilities and the conditional density functions are usually unknown. Furthermore, we often have no reason to assume that the prior probability is different for each of the decisions. In such cases we can use the expert rules and/or the learning set for the constructing decision algorithms.

The design of a decision tree structure in our approach to hierarchical classifier is based on human expert knowledge. In our consideration the decision rules are based on the probabilistic approach. The Bayes hierarchical classifier uses the Bayes theorem to design a classifier in each intermediate node.

The Bayesian hierarchical classifier contains a sequence of actions [2]. These actions are the simple classification tasks executed in individual nodes of the decision tree. Some specific features are measured on every level of the decision tree. At the first stage features x_0, at the second features x_1 are measured, and so on. Every set of features comes from the whole vector of features. In every node of the decision tree the classification is executed according to Bayes rule. The decisions i_1, i_2 … i_N are the results of recognition in the suitable node of the tree. At the last N-th stage, the decision made i_N indicates a single class. This class is the result of the Bayesian hierarchical classifier.

In our task of classification the number of classes is equal NC. The logic of making the decision is represented using the decision tree. The terminal nodes are labeled with the number of the classes from the M={1,2,...,NC}, where M is the set of labels classes. The non-terminal are labeled by numbers of 0, NC+1, NC+2… reserving 0 for the root-node.

Let's introduce the notation for the received model of multistage recognition [9]:

M(n) – the set of numbers of nodes, which distance from the root is n, $n = 0,1,2,...,N$.

In particular M(0)={0}, M(N)=M,

$\overline{\mathrm{M}} = \bigcup_{n=0}^{N-1} \mathrm{M}(n)$ – the set of interior node numbers (non-terminal),

$\mathrm{M}_i \subseteq \mathrm{M}(N)$ – the set of class numbers attainable from the i-th node ($i \in \overline{\mathrm{M}}$),

M^i – the set of numbers of immediate descendant nodes i ($i \in \overline{\mathrm{M}}$),

m_i – number of immediate predecessor of the i–th node ($i \neq 0$).

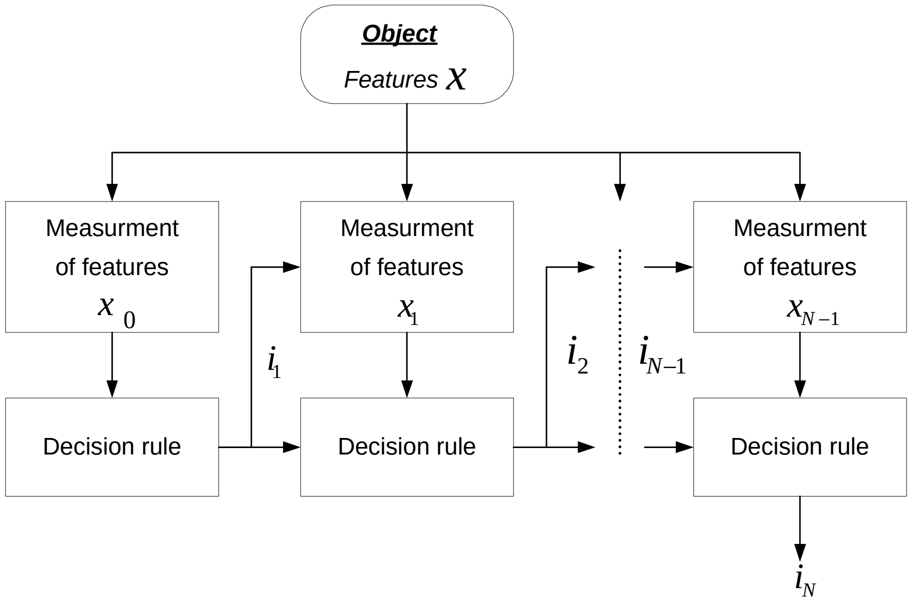

Fig. 1. Bayesian hierarchical classifier.

The Bayes hierarchical classifier is the example of the probabilistic model of pattern recognition. In this model the class of the pattern being recognised $j_N \in \mathrm{M}(N)$ is realization of random variable J_N and observed features x are realization of random variable X.

Our target now is to calculate the so-called multistage recognition strategy $\pi_N = \{\psi_i\}_{i \in \overline{\mathrm{M}}}$, that is the set of recognition algorithms in the form:

$$\Psi_i : \mathrm{X}_i \rightarrow \mathrm{M}^i, \; i \in \overline{\mathrm{M}} . \tag{5}$$

Presented formula is the decision rule (recognition algorithm) used at the i-th node which maps observation subspace to the set of immediate descendant nodes of the i-th node.

The strategy of the decision tree classifier represents the logic of making the decision. We favour two cases of the decision strategy. The first one is the locally optimal strategy. This strategy consists in minimizes the misclassification rate for particular nodes of a tree. Its decision rules are mutually independent. There are no relationships between nodes. The recognition algorithm at the n-th stage is as follows:

$$\overline{\Psi}_{i_n}(x_{i_n}) = i_{n+1} \;\; \text{when} \;\; \arg\max_{l \in \mathrm{M}^{i_n}} p(k) f_l(x_{i_n}) \; . \tag{6}$$

The second is globally optimal strategy. This strategy minimizes the mean probability of misclassification. The decision rules are mutually dependent by the empirical probability of correct classification. The recognition algorithm at the n-th stage is as follows:

$$\Psi^*_{i_n}(x_{i_n}) = i_{n+1} \;\; \text{when} \;\; \arg\max_{l \in \mathrm{M}^{i_n}} Pc(l) p(l) f_l(x_{i_n}), \tag{7}$$

where $Pc(l)$ is the empirical probability of correct classification at the next stages if at the n-th stage decision i_{n+1} is made.

As we mentioned bellow, in practice, the unknown probabilistic characteristics (values of the *prior* probabilities and probability density functions) are replaced by their estimators obtained via parametric or nonparametric approaches [5].

4. EXPERIMENTS

The first mathematical model of acute abdominal pain (APP) with decision tree was given in [8]. This model has sixteen classes and four stages of recognition. We simplified it to eight classes and two stages (see Fig. 2).

It leads to the following classification of the AAP:

1. cholecystitis,
2. pancreatitis,
3. non-specyfic abdominal pain,
4. rare disorders of “acute abdominal”,
5. appendicitis,
6. divercitulitis,
7. small-bowel obstruction,
8. perforated peptic ulcer.

The experts-physicians (from the Clinic of Surgery, Wroclaw Medical Academy) gave the decision tree presented in Fig.2. Numbers of leafs are the numbers of diagnosis and the numbers in the nodes are correspond to the following diagnoses:

9. acute enteropathy,
10. acute disorders of the digestive system,
11. others.

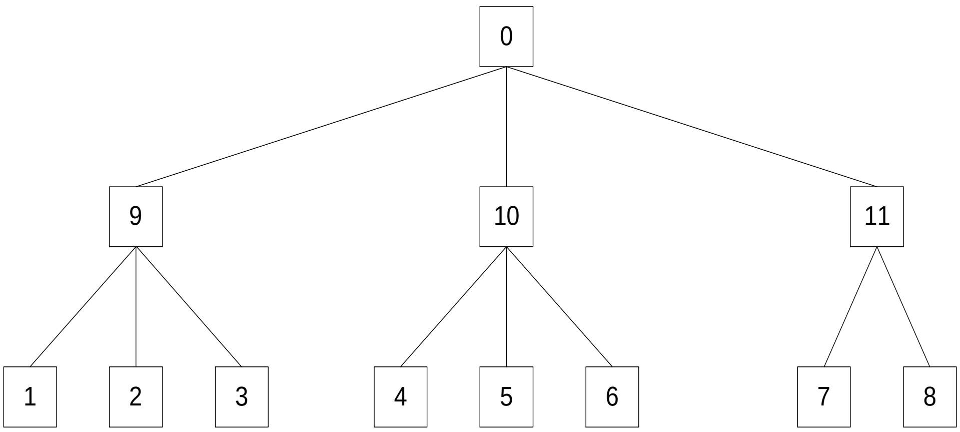

Fig. 2. Heuristic classifier for the APP diagnosis problem.

5. EXPERIMENTAL INVESTIGATION

In the experiment we used the multistage classifiers with heuristic decision tree and Bayesian classifier in each node according the local optimal strategy. For these classifiers the estimators of the conditional probability density function were obtained via k-Nearest Neighbor [4]. Dataset used in experiment was collected in Clinic of Surgery, Wroclaw Medical Academy and consists of 476 clinical histories. The clinical feature description is presented in Tab.1. The errors of the classifiers were estimated using the ten fold cross validation method [7]. Features used in each node of decision tree are presented in Tab. 2. Diagnostic accuracy (DA) was defined by the number of correct predictions divided by the total number of elements in dataset.

The aim of the experiment is to compare the errors for different values of k. The selection of features has been made in accordance with the suggestions of other work on this topic [1,8].

Table 1. Clinical feature description.

no	Attribute	no	Attribute	no	Attribute
1	Sex	12	nausea and vomiting	23	Pulse
2	Age	13	Appetite	24	respiratory movements of abdomen
3	pain location on the beginning	14	bowel movement	25	Flatulence
4	pain location on present	15	Urinate	26	Tenderness (location)
5	pain intensity	16	previous indigestion	27	Blumberg's sign
6	aggravating factors	**17**	Jaundice	28	muscle's Demence
7	relieving factors	18	previous burgery (abdominal)	29	increased tension of abdominal
8	pain progression	19	Drugs	30	Swellings
9	pain duration	20	Mood	31	Murphy's sign
10	pain type on the beginning	21	skin's color		
11	pain type at present	**22**	Temperature		

Table 2. Features used in nodes of decision tree.

node	Features						
0	26	30	11	6	16	31	10
9	26	3	19	4	31	16	10
10	27	3	11	6	28	15	26
11	27	11	29	6	26	10	23

The values of error rate for node "0" are presented in Tab. 3. Tab.4-8 show the values of error rate on the second stage of classification.

Table 3. Error rate on the first stage of classification.

k	1	3	5	7	9
"0"	0,313	0,288	**0,275**	0,286	0,286

Table 4. Error rate on the second stage of classification on condition that on first stage k = 1.

k	1	3	5	7	9
"9"	0,171	0,136	0,096	**0,096**	0,104
"10"	0,229	0,253	0,193	0,205	0,193
"11"	0	0	0	0	0
DA	0,435	0,428	0,397	0,400	0,400

Table 5. Error rate on the second stage of classification on condition that on first stage k = 3.

k	1	3	5	7	9
"9"	0,222	0,163	0,126	0,119	0,141
"10"	0,221	0,221	0,182	0,182	0,208
"11"	0	0	0	0	0
DA	0,428	0,405	0,381	0,378	0,393

Table 6. Error rate on the second stage of classification on condition that on first stage k = 5.

k	1	3	5	7	9
“9”	0,193	0,179	0,131	0,117	0,131
“10”	0,257	0,171	0,157	0,157	0,171
“11”	0	0	0	0	0
DA	0,413	0,387	0,365	0,359	0,368

Table 7. Error rate on the second stage of classification on condition that on first stage k = 7.

k	1	3	5	7	9
“9”	0,186	0,172	0,152	0,138	0,124
“10”	0,231	0,200	0,169	0,185	0,200
“11”	0	0	0	0	0
DA	0,415	0,402	0,386	0,384	0,383

Table 8. Error rate in the second stage of classification on condition that on first stage k = 9.

k	1	3	5	7	9
“9”	0,162	0,176	0,142	0,122	0,135
“10”	0,170	**0,102**	0,136	0,102	0,102
“11”	0	0	0	0	0
DA	0,390	0,379	0,374	0,358	0,363

Table 9. Error rate for one stage classification for features set {26, 30, 11, 6, 16, 31, 10}.

k	1	3	5	7	9
DA	0,399	0,345	0,345	0,374	0,394

6. DISCUSION

The following conclusions could be drawn from the experiment:

At node 11 we deal with error-free classification. Incorrect diagnosis of “small-bowel obstruction” or “perforated peptic ulcer” occurs only during the misclassifications in the node 0.

In this decision tree different classifiers (parameters) in each node should be used. This approach can increase significantly the quality of classification. In our research the difference between best and worst result is equal 7,7% of diagnostic accuracy.

In the comparison with standard k-NN classifier we received slightly worse results. Further research may relate to changes in the structure of decision tree and use of other classifiers. It seems that modifying the tree structure can further improve the quality of classification.

7. CONCLUSION

The recognition methods based on hierarchical Bayesian approach was presented. The classifier based on Bayes rule was applied to the medical decision problem (recognition of Acute Abdominal Pain) and can be used to help the clinicians to make their own diagnosis.

The presented heuristic results for choice of parameter k demonstrate the effectiveness of the proposed concepts in such computer-aided medical diagnosis problems. Further research may relate to changes in the structure of decision tree. It seems that modifying the tree structure can further improve the quality of classification.

8. ACKNOWLEDGMENTS

This work is supported in part by the National Science Centre under a grant for the period 2011-2014.

BIBLIOGRAPHY

[1] BURDUK R., WOŹNIAK M., Bayes Multistage Classifier and Boosted C4.5 Algorithm in Acute Abdominal Pain Diagnosis, Advances in Intelligent and Soft Computing, Vol. 59, 2009, pp. 371-378.
[2] BURDUK R., KURZYŃSKI M., Two-stage binary classifier with fuzzy-valued loss function, Pattern Analysis and Applications, Vol. 9, No 4, 2006, pp. 353-358.
[3] DE DOMBAL F.T., LEAPER D.J., STANILAND J.R., McCANN A.P., HORROCKS C., Computer-aided diagnosis of acute abdominal pain, Br. Med. J. II, 1972, pp. 9-13.
[4] DEVIJVER P.A., KITTLER J,. Pattern Recognition: A Statistical Approach, Prentice Hall, London, 1982.
[5] DUDA R.O., HART P.E., STORK D.G., Pattern Classification, John Wiley and Sons, 2000.
[6] EICH H.P., OHMANN C., LANG K., Decision support in acute abdominal pain using an expert system for different knowledge bases, Proceedings of the 10th IEEE Symposium on Computer-Based Medical Systems, 1997, pp. 2-7.
[7] KOHAVI R., A study of cross-validation and bootstrap for accuracy estimation and model selection, Proceedings of the 14th International Joint Conference on Artificial Intelligence, San Mateo, 1995, pp. 1137–1143.
[8] KURZYŃSKI M., Diagnosis of acute abdominal pain using three-stage classifier, Computers in Biology and Medicine, Vol. 17, No 1, 1987, pp. 19-27.
[9] KURZYŃSKI M., On the Multistage Bayes Classifier, Pattern Recognition, Vol. 21, 1998, pp. 355-365.
[10] MUI J., FU K.S., Automated classification of nucleated blood cells using a binary tree classifier, IEEE Trans. Pattern Anal. Mach. Intell. Vol. PAMI-2, 1980, pp. 429-443.
[11] OHMANN C., MOUSTAKIS V., YANG Q., LANG K., Evaluation of automatic knowledge acquisition techniques in the diagnosis of acute abdominal pain, Artif Intell Med., 1996, Vol. 8, No. 1, pp. 23-36.
[12] SAFAVIAN, S.R., LANDGREBE, D., A survey of decision tree classifier methodology, IEEE Trans. Systems, Man Cyber, 21 (3), 1991, pp. 660-674.