Proposition of Boosting Algorithm for Probabilistic Decision Support System

Michal Wozniak

Chair of Systems and Computer Networks, Wroclaw University of Technology,
Wybrzeze Wyspianskiego 27, 50-370 Wroclaw, Poland
Michal.Wozniak@ pwr.wroc.pl

Abstract. Different experts formulate the rules with different qualities. Additional we may get some information about problem from databases and qualities of information stored in the databases are different. We will propose the quality measure of knowledge we got. We will show how use it for decision process based on Bayes formulae and boosting concept.

1 Introduction

During designing decision support systems we get the rules from different sources (experts, databases) and their qualities are different. The following paper concerns on the decision making on the base on the different classifiers through voting procedure. This concept called *boosting* [4] will be used to the probabilistic decision making. The organization of voting system is based on the qualities of information sources.

The content of the work is as follow: Next section presents proposition of statistical knowledge quality measure and it shows how use proposed quality measure for boosting decision making. In section 3 the results of experimental investigation of proposed decision method are presented. The last section concluded the paper.

2 Boosting Concept for Probabilistic Reasoning

For the knowledge given by experts we can not assume that expert tell us true or the rule set is generated (by the machine learning algorithms) on the noise-free learning set. We postulate that we believe on it only with the γ factor ($P(\text{rule})=\gamma\leq 1$), proposed as the confidence (quality) measure[6]. For the practical cases the value of proposed measure is constant for each rule obtained from the same expert or generated on the base on the same learning set. Therefore let $\gamma^{(K)}$ denotes confidence measure of *K*-th source of knowledge.

The Bayes decision theory consists of assumption [1] that the feature vector x and number of class j are the realization of the pair of the random variables X, J. The formalisation of the recognition in the case under consideration implies the setting of an optimal Bayes decision algorithm $\Psi(x)$, which minimizes probability of misclassification for 0-1 loss function:

M. Bubak et al. (Eds.): ICCS 2004, LNCS 3036, pp. 675–678, 2004.
© Springer-Verlag Berlin Heidelberg 2004

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

In the real situation the *posterior* probabilities for each classes are usually unknown. Instead of them we can used the rules and/or the learning set for the constructing decision algorithms[5].

The analysis of different practical examples leads to the following form of rule $r_i^{(k)}$:

IF $x\in D_i^{(k)}$ **THEN** state of object is *i* **WITH** posterior probability $\beta_j^{(k)}=\int_{D_i^{(k)}} p(i|x)dx$

greater than $\underline{\beta}_i^{(k)}$ and less than $\overline{\beta}_i^{(k)}$.

Lets note the rule estimator will be more precise if rule decision region and differences between upper and lower bound of the probability given by expert will be smaller. For the logical knowledge representation the rule with the small decision area can be overfitting the training data [2]. For our proposition we respect this danger for the rule set obtained from learning data. For the estimation of the *posterior* probability from rule we assume the constant value of for the rule decision area. Therefore lets propose the relation "more specific" between the probabilistic rules pointed at the same class.

Definition. Rule $r_i^{(k)}$ is "more specific" than rule $r_i^{(l)}$ if

$$\left(\overline{\beta}_i^{(k)}-\underline{\beta}_i^{(k)}\right)\left(\int_{D_i^{(k)}} dx \Big/ \int_X dx\right) < \left(\overline{\beta}_i^{(l)}-\underline{\beta}_i^{(l)}\right)\left(\int_{D_i^{(l)}} dx \Big/ \int_X dx\right) \tag{2}$$

Hence the proposition of the *posterior* probability estimator $\hat{p}(i|x)$ is as follow:

`from subset of rules` $R_i(x)=\left\{r_i^{(k)}: x\in D_i^{(k)}\right\}$ `choose the "most specific" rule` $r_i^{(m)}$

$$\hat{p}(i|k)=\left(\overline{\beta}_i^{(m)}-\underline{\beta}_i^{(m)}\right)\Big/\int_{D_i^{(m)}} dx \tag{3}$$

When only the set *S* is given, the obvious and conceptually simple method is to estimate *posterior* probabilities $\hat{p}(i|x)$ for each classes via estimation of unknown conditional probability density functions (CPDFs) and *prior* probabilities.

For the considered case, i.e. when some of rule sets and learning set are given, we propose the boosting probabilistic algorithm $\psi^{(B)}(x)$:

$$\psi^{(B)}(x)=i \text{ if } p^{(B)}(i|x)=\max_{k\in M} p^{(B)}(k|x), \tag{4}$$

$$p^{(B)}(i|x)=\sum_{K=1}^{N}\gamma^{(K)}\hat{p}(i|x)\Big/\sum_{K=1}^{N}\gamma^{(K)}. \tag{5}$$

N denotes number of knowledge source (experts and learning sets), $\hat{p}(i|x)$ denotes estimator of *posterior* probability obtained on base on the learning set or rule one.

3 Experimental Investigations

In order to appreciate the proposed concept several experiments were made on the computer-generated data. We have restricted our considerations to the case of rules for whose the upper and lower bounds of the *posterior* probabilities are the same, rule defined region for each $i \in \{1, ..., M\}$ cover the whole feature space X and we have only one learning set. In experiments our choice of the CPDFsand the *prior* probabilities was deliberate. In experiments we considered a two-class recognition task with set of 6 and 10 rules, the Gaussian CPDFs of scalar feature x and with following parameters $p_1 = 0.333$, $p_2 = 0.667$, $f_1(x) = N(0,1), f_2(x) = N(2,1)$.

The value of quality measure of learning set has been counted using following heuristic formulae $\gamma^{(s)} = \text{size of learning set} * 0{,}001$. The value of quality measure of rule has been counted using heuristic formulae $\gamma^{(R)} = \text{number of rules} * 0{,}1$. $D^{(k)}$ denotes decision area of the rule, $\beta_1^{(k)}$ denotes *posterior* probability estimator of rule pointed at class 1, $\beta_2^{(k)}$ denotes *posterior* probability estimator of rule pointed at class 2.

Table 1. Rule sets for experiments

Experiment 1				Experiment 2			
k	$D^{(k)}$	$\beta_1^{(k)}$	$\beta_2^{(k)}$	k	$D^{(k)}$	$\beta_1^{(k)}$	$\beta_2^{(k)}$
1	[-3,0, -1,0)	0,984	0,016	1	[-3,0, -2,0)	0,997	0,003
2	[-1,0, 0,0)	0,887	0,113	2	[-2,0, -1,0)	0,982	0,018
3	[0,0, 1,0)	0,556	0,444	3	[-1,0, 0,0)	0,887	0,113
4	[1,0, 2,0)	0,166	0,834	4	[0,0, 0,5)	0,684	0,316
5	[2,0, 3,0)	0,031	0,969	5	[0,5, 1,0)	0,449	0,551
6	(3,0, 5,0]	0,004	0,996	6	(1,0, 1,5]	0,234	0,766
				7	(1,5, 2,0]	0,103	0,897
				8	(2,0, 3,0]	0,031	0,969
				9	(3,0, 4,0]	0,005	0,995
				10	(4,0, 5,0]	0,001	0,999

The results of experiments are shown on the Fig. 1.
The following conclusion may be drawn from the experiments:

- The frequency of correct classification depends on the value of the confidential measure of rules. The algorithms with bigger value always give the better results.
- Boosting algorithms lead to better or similar results compared to algorithm k-NN, especially for the big learning set.

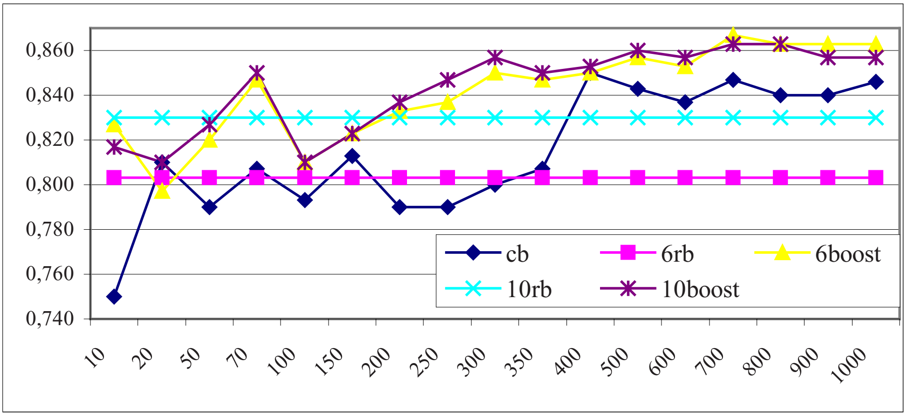

Fig. 1. Frequency of correct classification for the experiments. Cb denotes case-based algorithm, 6rb rule-based one which used 6 rules, 10rb rule based one with 10 rules, 6boost boosting algorithm for 6rb and cb, 10boost boosting one for 10rb and cd.

Drawing a general conclusion from such a limited scope of experiments as described above is of course risky. However results of experimental investigations encourage applying proposed algorithms in practise.

4 Conclusion

The paper concerned probabilistic reasoning and the proposition of the quality measure for that formulated decision problems. We presented how use the proposed measure in decision algorithm based on boosting concept.
Presented ideas need the analytical and simulation researches but the preliminary results of the experimental investigations are very promising.

References

1. Duda R.O., Hart P.E., Stork D.G., *Pattern Classification*, Wiley-Interscience, 2000.
2. Mitchell T., *Machine Learning*, McGraw Hill, 1997.
3. Puchala E., A Bayes Algorithm for the Multitask Pattern Recognition Problem – Direct Approach, *LNCS no 2659,* 2003.
4. Schapire R. E., The boosting approach to machine learning: An overview. *Proc. Of MSRI Workshop on Nonlinear Estimation and Classification*, Berkeley, CA, 2001.
5. Walkowiak K., A Branch and Bound Algorithm for Primary Routes Assignment in Survivable Connection Oriented Networks, *Computational Optimization and Applications*, Kluwer Academic Publishers, February 2004, Vol. 27.
6. Wozniak M., Concept of the Knowledge Quality Management for Rule-Based Decision System, W: Klopotek M.A. et al. [eds] *Intelligent Information Processing and Web Mining*, Springer Verlag 2003.