Application of the Confidence Measure in Knowledge Acquisition Process

Michal Wozniak

Chair of Systems and Computer Networks, Wroclaw University of Technology, Poland The Higher Vocational State School in Legnica, Poland wozniak@zssk.pwr.wroc.pl

Abstract. Paper deals with the knowledge acquisition process. Different experts formulate the rules for decision support systems. We assume they have different knowledge about the problem and therefore obtained rules have different qualities. We will formulate the proposition of the confidence measure and its application to the decision process based on Bayes formulae. We will propose how calculate the value of measure under consideration for typical statistical learning process. On the base on the proposed measure of the knowledge quality we propose the procedure of the contradictions elimination for the set of logical rules.

1 Introduction

Machine learning is the attractive approach for building decision support systems [9]. For this type of software, the quality of the knowledge base plays the key-role. In many cases we can meet following problems:

- the experts can not formulate the rules for decision problem, because they might not have the knowledge needed to develop effective algorithms, but we can get the learning data from databases and experts only classify each record to the correct class,
- the knowledge given by experts is uncompleted or qualities of experts (knowledge sources) are different for each of them.

For the first problem machine learning algorithms can give us the set of rules (hypothesis) on the base on the learning set, but we have to answer following questions:

- 1. Who made the object descriptions? (Can we trust the operator?)
- 2. Who confirmed the diagnosis and what was the expert quality?

In the second problem we get the rules from different experts and their qualities are different. This problems was partly described for the induction learning [2,3,7] and for the concept description [1]. The following paper concerns on the quality of rule for the probabilistic reasoning, but the proposed measure can be modified to acquisition process for another form of rule.

The content of the work is as follow: Section 2 introduces necessary background and provides the probabilistic decision problem statement. Next section presents the form of rule for the probabilistic expert systems and proposes the rule-based algorithm. Section 4 defines statistical confidence measure of the knowledge and shows how modify knowledge base according to confidence measure of rules. In this section we also presents the interpretation of proposed measure for the estimation process based on the typical statistical model. Section 5 proposes the procedure of contradictions elimination for the set of rules obtained from different sources. The last section concluded the paper.

2 Decision Problem Statement

Among the different concepts and methods of using "uncertain" information in pattern recognition, an attractive from the theoretical point of view and efficient approach is through the Bayes decision theory. This approach consists of assumption [5] that the feature vector $x = (x^{(1)}, x^{(2)}, ..., x^{(d)})$ (describing the object being under recognition) and number of class $j \in \{1, 2, ..., M\}$ (the object belonged to) are the realization of the pair of the random variables X, J. For example in medical diagnosis X describes the result of patient examinations and J denotes the patient state. Random variable J is described by the prior probability p_i , where

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

X has probability density function

$$f(X = x | J = j) = f_j(x)$$
⁽²⁾

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

$$p(j|x) = \frac{p_{j}f_{j}(x)}{\sum_{k=1}^{M} p_{j}f_{j}(x)}$$
(3)

The formalisation of the recognition in the case under consideration implies the setting of a optimal Bayes decision algorithm $\Psi(x)$, which minimizes probability of misclassification for 0-1 loss function[4]:

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

In the real situation the prior probabilities and the conditional density functions are usually unknown. Furthermore we often have no reason to decide that the prior probability is different for each of the decisions. Instead of them we can used the rules and/or the learning set for the constructing decision algorithms [12,13].

3 Rule-Based Decision Algorithm

Rules as the form of learning information is the most popular model for the logical decision support systems. For systems we consider the rules given by experts have rather the statistical interpretation than logical one.

The form of rule for the probabilistic decision support system[6] is usually as follow

if A then B with the probability
$$eta_{i}$$

where β is interpreted as the estimator of the *posterior* probability, given by the following formulae:

$$\beta = P(B|A) \tag{5}$$

More precisely, in the case of the human knowledge acquisition process, experts are not disposed to formulate the exact value of the β , but he (or she) rather prefers to give the interval for its value

$$\beta \le \beta \le \overline{\beta} \ . \tag{6}$$

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

$$\beta_j^{(k)} = \int_{D_i^{(k)}} p(i|x) dx \tag{7}$$

The graphical interpretation of the *posterior* probability estimator for the decision area given by rule is depicted on Fig. 1.



Fig. 1. Graphical interpretation of the posterior probability estimator given by the expert rule

For that form of knowledge we can formulate the decision algorithm $\Psi_R(x)$

$$\Psi_R(x) = i \quad \text{if} \quad \hat{p}(i|x) = \max_k \, \hat{p}_k\left(i|x\right),\tag{8}$$

where $\hat{p}(i|x)$ is the *posterior* probability estimator obtained from the rule set.

The knowledge about probabilities given by expert estimates the average *posterior* probability for the whole decision area. As we see for decision making we are interested in the exact value of the *posterior* probability for given observation.

Lets note the rule estimator will be more precise if:

- rule decision region will be smaller,
- differences between upper and lower bound of the probability given by expert will be smaller.

For the problem under consideration definition of the relation "more specific" between the probabilistic rules pointed at the same classes is very useful.

Definition

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

$$\left(\overline{\beta}_{i}^{(k)} - \underline{\beta}_{i}^{(k)}\right) \left(\frac{\int dx}{\int dx}_{X}\right) < \left(\overline{\beta}_{i}^{(l)} - \underline{\beta}_{i}^{(l)}\right) \left(\frac{\int dx}{\int dx}_{X}\right).$$
⁽⁹⁾

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 $r_i^{(m)}$.

$$\hat{p}(i|k) = \frac{\left(\overline{\beta}_{i}^{(m)} - \underline{\beta}_{i}^{(m)}\right)}{\int\limits_{D_{i}^{(m)}} dx}$$
(10)

4 Proposition of Knowledge Confidence Measure

4.1 Definition

This estimator $\hat{p}(i|x)$ is obtained under the following assumption:

- learning set is noise free (or expert tell us always true),
- target concept contained in the set of class number $\{1, ..., M\}$,
- the prior probabilities of classes are unknown.

We consider decision under the first assumption given by the following formulae: $p(If A then B with methability \beta) = 1$

 $P(\text{If A then B with probability }\beta) = 1.$

During the expert system designing process the rules are obtained from different sources and the sources have the different confidence. For the knowledge given by experts we can not assume that expert tell us true or/and if the rule set is generated by the machine learning algorithms we can not assume the learning set is noise free.

Therefore we postulate that we have not to trust all information we got or the believe on it only with the γ factor, proposed as the confidence (quality) measure. It can be formulated as [14]

P(If A then B with probability $\beta) = \gamma \le 1$.

Lets as $\gamma_i^{(k)}$ denote the value of the confidence measure of rule $r_i^{(k)}$.

4.2 Using Confidence Measure for Rule Base Modification

We defined confidence measure. Now, lets us show how it can be utilized for the modification the set of rules. For the form of rule we described in section 3 we propose the following procedure, which should be started after the acquisition process:

for i:=1 to M
for each rule
$$r_i^{(k)}$$
 pointed at class i:
 $\overline{\beta}_i^{(k)} = 1 - (1 - \overline{\beta}_i^{(k)}) \gamma_i^{(k)};$
 $\underline{\beta}_i^{(k)} = \underline{\beta}_i^{(k)} \gamma_i^{(k)}$
endfor.

endfor.

The proposition of the application of the confidence measure for the set of the logical rules will be shown in the next section.

4.3 Confidence Measure for the Statistical Estimation

The central problem of our proposition is how to calculate the confidence measure. For human experts the values for their rules is fixed arbitrary according to the quality of creator. The presented problem we can also find in the typical statistical estimation of unknown parameter β , where we assume the significant level[12]. The significant level can be interpreted as the confidence measure. Each rule gives the index of the class. If the feature vector value belongs to the decision area given by the rule, the decision depends on the previous state and on the applied therapy. While constructing the artificial rule set, we have to define somehow the decision areas for the new rule set. For example we can want to obtain for each rule *posterior* probability estimator, which is not less than a fixed value or in the practice we can use the one of very well known machine learning algorithms like AQ,CN2 [9].

For each of the given intervals k we have to obtain the estimator of the *posterior* probability.

We use the following statistical model [11]:

- the learning set is selected randomly from a population and there exists two class of points: marked (point at the class $i \in \{1, ..., M\}$) and unmarked (point at the class *l*, where $l \in \{1, ..., M\}$ and $l \neq i$),
- the expected value for the population is p,
- the best estimator of p is

$$\hat{p} = \frac{m}{n},\tag{11}$$

where n means the sample size and m - the number of the marked elements.

Let us concentrate on two cases

Small Sample ($n \le 100$)

$$t = \frac{\hat{p} - p}{s/\sqrt{n}} \tag{12}$$

has the Student's *t*-distribution. We want to estimate one parameter - the expected value, therefore we use *t*-distribution with n-1 degree of freedom.

For the fixed significance level α we get

$$P(t| < \mu_{\alpha}) = 1 - \alpha \tag{13}$$

using the short-cut formula

$$s^{2} = \frac{\sum (x_{i} - \hat{p})^{2}}{(n-1)} \approx \sum x_{i}^{2} + \frac{(\sum x_{i})^{2}}{n}$$
(14)

we obtain

$$P\left(\frac{m}{n} - \mu_{\alpha}\sqrt{\frac{m\left(1 - \frac{m}{n}\right)}{n}}
$$(15)$$$$

The μ_{α} is the value of the *t*-distribution on *n*-1 degrees of freedom and for the significance level α . In this case we get rule $r_i^{(k)}$, for which confidence measure of rule $\gamma_i^{(k)}$ is given by the following equation

$$\gamma_i^{(k)} = 1 - \alpha , \qquad (16)$$

$$\underline{\beta}_{i}^{(k)} = \frac{m}{n} - \mu_{\alpha} \sqrt{\frac{\frac{m}{n} \left(1 - \frac{m}{n}\right)}{n}} \text{ and } \overline{\beta}_{i}^{(k)} = \frac{m}{n} + \mu_{\alpha} \sqrt{\frac{\frac{m}{n} \left(1 - \frac{m}{n}\right)}{n}}$$
(17)

Big Sample (*n* > 100)

For the big sample, the distribution is similar to the normal distribution. We thus have the equation of the same for as (15), but in this case μ_{α} is the value of normal standardized N(0, 1) distribution for the significance level α .

5 Contradictions Elimination Algorithm

As we have mentioned the proposed method of the quality management can be applied to the logical knowledge representation (where "if-then" means logical implication). E.g. for the unordered set [2,9] of logical rules acquisition process we can attribute the value of confidence to each of rule. It could be used in the case if the contradiction in the set of rules would be detected.

First we note the set of rule R consists of the M subsets

$$R = R_1 \cup R_2 \cup \dots \cup R_M , \qquad (18)$$

where R_i denote subset of rule pointed at the *i*-th class.

For this form of rule the two of them contradict each other if

$$\exists x \in X \land \exists k, l \in \{1, 2, ..., M\}, k \neq l \land \exists i, j$$

$$x \in D_i^{(k)} \land x \in D_j^{(l)}$$
(19)

where *i*, j denote the number of rule.

The equation (19) means that we can find observation, which belongs to the decision area of the rule pointed at class $i(R_i^{(k)})$ and decision areas of the rule pointed at different class $j(R_i^{(l)})$.

Let us propose the idea of the contradictions elimination algorithm.

//for each class number
for i:= 1 to M:
//for each rule in
$$R_i$$

for k:=1 to $|R_i|$
//for each class number bigger than *i*
for j:= i to M:
//for each rule in R_j
for l:= 1 to $|R_j|$:
//if $r_i^{(k)}$ and $r_j^{(l)}$ contradict each other

if $D_i^{(k)} \cap D_i^{(l)} \neq \emptyset$ then //if confidence of $r_i^{\left(k\right)}$ is higher then //confidence of $r_i^{(l)}$ if $\gamma_i^{(k)} \ge \gamma_i^{(l)}$ then //remove the common part from decision //area of rule $r_i^{(l)}$ $D_i^{(l)} \coloneqq D_j^{(l)} \setminus \left(D_j^{(l)} \cap D_i^{(k)} \right)$ else //remove the common part from decision //area of rule $r_i^{(l)}$ $D_i^{(k)} \coloneqq D_i^{(k)} \setminus \left(D_i^{(l)} \cap D_i^{(k)} \right)$ fi fi endfor endfor endfor endfor

6 Conclusion

The paper concerned probabilistic reasoning and the proposition of the quality measure for that formulated decision problems. We proposed the idea of the contradiction detection and elimination method for the logical representation of experts' knowledge too. We hope this idea of confidence management can be helpful for other problems whose can be met during the knowledge acquisition process from different sources.

Presented ideas need the analytical and simulation researches. Let us draw some future works under the concept of the information quality:

- 1. developing the method how to judge the expert quality (we formulated only the method of counting the confidence measure for rules obtained via machine learning algorithms but we propose arbitrary judgment for rule given by experts),
- 2. applying proposed method to the real medical decision problems (work in progress),
- 3. analytical researches into proposed method properties,
- performing simulation experiments on computer generated data to estimate the dependencies between the size of the decision area and the data quality versus correctness of classification,
- 5. developing the software for the contradictions elimination algorithm.

References

- Bergadano F., Matwin S., Michalski R.S., Zhang J., Measuring of Quality Concept Descriptions, *Proc. Of the 3rd European Working Session on Learning*, Aberdeen, Scotland, 1988.
- 2. Bruha I., Kockova S., Quality of decision rules: Empirical and statistical approaches, *Informatica*, no 17.
- 3. Dean P., Famili A., Comparative Performance of Rule Quality Measures in an Inductive Systems, *Applied Intelligence*, no 7, 1997.
- 4. Devijver P.A., Kittler J., *Pattern Recognition: A Statistical Approach*, Prentice Hall, London 1982.
- 5. Duda R.O., Hart P.E., *Pattern Classification and Scene Analysis*, John Wiley and Sons, New York, 1973
- 6. Giakoumakis E., Papakonstantiou G., Skordalakis E., Rule-based systems and pattern recognition, *Pattern Recognition Letters*, No 5, 1987.
- Gur-Ali O., Wallance W.A., Induction of rules subject to a quality constraint: probabilistic inductive learning, *IEEE Transaction on Knowledge and Data Engeineering*, vol. 5, no 3, 1993.
- 8. Kurzynski M., Wozniak M.: Rule-based algorithms with learning for sequential recognition problem, *Proceedings of the Third International Conference on Information Fusion "Fusion 2000"*. Paris, July 10–13, 2000.
- 9. Mitchell T., Machine Learning, McGraw Hill, 1997.
- 10. Pearl J., *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann Pub. Inc., San Francisco, California, 1991.
- 11. Sachs L., *Applied Statistic. A Handbook of Techniques*, Springer-Verlag, New York Berlin Heideberg Tokyo, 1984.
- 12. Wozniak M., Blinowska A., Unification of the information as the way of recognition the controlled Markov chains, *Proc. of the Congress on Information Processing and Management of Uncertainty in Knowledge Based Systems*, Granada, Spain 1996.
- Wozniak M., Jackowski K., Rule-based contextual pattern recognition algorithm parametric case. *Proceedings of International Symposium on Pattern Recognition*, Royal Military Academy. Brussels, February 12, 1999.
- 14. Wozniak M., Idea of Knowledge Acquisition for the Probabilistic Expert Systems, *Proceedings of the International Conference on Computational Intelligence for Modeling, Control and Automation*, Vienna (Austria), 12–14 February 2003 (be published).