USING VALUED CLOSENESS RELATION IN CLASSIFICATION SUPPORT OF NEW OBJECTS

Jerzy Stefanowski
Institute of Computing Science,
Poznań University of Technology,
PL-60-965 Poznań, Poland
e-mail: stefanj@pozn1v.tup.edu.pl

Key words: Machine learning, Classification systems

Abstract

The problem of using the decision rule sets in the classification support of new objects is being addressed. The important part of this support is solving cases where the new object matches condition parts of the decision rules in an ambiguous way or it does not match any of them. The valued closeness relation is used in such cases. Several computation experiments on different real-life data sets were performed to evaluate the usefulness of the valued closeness concept. The obtained results are presented and discussed in the paper.

1 Introduction

Classifying is one of the main activities in the decision making. The problems of supporting decisions concerning classification of objects and its experimental verification are being considered in this paper.

The classification support is based on the knowledge derived from the set of learning examples. The learning examples are represented in an *attribute-value* form and refer to the examples of real or hypothetical decisions about classification of objects (persons, observations, processes, etc.) on the basis of their description by attributes (features, characteristics, etc.).

Learning from examples is an area of interest of many methods coming from different fields, i.e. generally from machine learning, neural networks or statistics (cf. Weiss and Kulikowski 1991). In the following paper, knowledge about classification is represented in the form of *decision rules*. We assume, similarly as in many papers in the area of machine learning, that rules are in the following form:

$$if \ \ (a_1, v_1) \,\&\, (a_2, v_2) \,\&\ldots\&\, (a_n, v_n) \ \ then \ (d, v_d)$$

where a_i is i-th condition attribute, v_i is a value of this attribute, $i = 1, 2, \ldots, n$; d is a classification decision and v_d is its value; objects having the same value v_d of decision attribute d create the so called decision class.

The decision rules can be induced from learning examples in several ways. Although the concept of classification supporting is general, it will be restricted in this paper to the rule sets obtained by methods inducing the so called *minimal rule description*. These methods are focused on generating in a heuristic way a sufficient number of the most general rules which cover given decision classes.

Learning systems are often faced with inconsistent data. Examples are inconsistent if they have the same description (by condition attribute values) and belong to different decision classes. One of the best approaches to handle such inconsistencies is the *rough set theory*, introduced by Z. Pawlak (Pawlak 1991).

If the set of learning examples is inconsistent then *lower* and *upper approximations* of each decision class are computed. It allows us to obtain two kinds of rules: *exact* and *approximate*. Exact rules are induced from examples belonging to lower approximation of decision class, while approximate rules are induced from examples belonging to boundaries or upper approximations. The approximate rules can indicate several classification decisions possible for given conditions.

In the following paper, the decision rules are induced using the author's implementation of *LEM* algorithm based on the idea of *single local covering* introduced in (Grzymala 1992). In this implementation, unlike the original version of *LEM* algorithm, the approximate rules are generated from the boundaries of decision classes.

The set of decision rules is the basis of classification support for *new objects*. By new objects we understand objects unseen in the learning phase. The support is performed by matching a new object description to condition parts of decision rules.

One of the possible results of such a matching is the situation where a new object does not match any of the rules. In this case one can use rules "*nearest*" to the description of the new object. The *valued close-*

ness relation is used to look for the nearest rules (cf. Slowinski and Stefanowski 1993). This relation is constructed by using preference information interactively acquired from the decision maker.

The presented idea of using valued closeness relation in classification supporting has been implemented as an interactive software system called $VCR-Class$.

The main aim of this paper is to present the results of experimental evaluation of this system. This evaluation is based on several computation experiments where the correctness of classification decisions suggested by the system have been checked. Different real-life data sets were used for experiments.

2 Classification support using nearest rules

The sets of decision rule are used to classify new objects. In general, it is done by matching the classified object description to condition parts of induced rules. During classification of the object three possible cases may happen:

(a) the new object matches exactly one rule,
(b) the new object matches more than one rule,
(c) the new object does not match any of the rules.

In case **(a)**, if the matched rule is an exact one then the classification suggestion is clear. In case of matching to approximate rule the suggestion may be ambiguous. Similar difficulties occur in determining suggestions for case **(b)**.

One should notice, however, that in some rule classification systems not all of the above cases may happen. In a rule option of the system $C4.5$ (Quinlan 1994) and in the first version of the $CN2$ system (Clark and Niblett 1989) rules are ordered. The matching is done starting from the top of the ordered list till the first matched rule is found. In case of no matching, the default rule (which is extra added to the end of the rule list) is used.

In the classification system proposed in this paper less arbitrary solution has been chosen. During classification, the whole rule set is scanned (like in classical Michalski's AQ system (Michalski *et al* 1986)). Analysis of ambiguous cases (i.e. **(b)** or **(a)** – matching to an approximate rule) is done by using additional information about learning examples supporting matched rules. By *supporting examples* we understand examples satisfying the condition part of a given rule. For each decision rule a coefficient of its strength is determined. The *strength* of a rule is the number of supporting examples. Thus, in an ambiguous case **(a)** the decision maker is informed about the strength of each possible decision. If the strength of one decision is significantly greater than others, the decision maker may conclude that the new object most likely belongs to this decision. Case **(b)** is solved in a similar way. Information about all matched rules is extended by their strength and the decision maker may choose the strongest possible decision.

Case **(c)** is the most difficult to interpret. In this case, one can help the decision maker by presenting him a set of rules "*nearest*" to the description of the new object. The nearest rules are rules which are close to the description of a new object in the sense of a certain *distance measure*. In other words, they do not differ from the classified object in a significant way (usualy they are partly matched with the object).

One can also notice that other, a bit different approaches to solve this case have been already introduced. The first one was proposed by Michalski in his $AQ15$ system (Michalski 1986). Other proposals have been also presented (Stefanowski 1993), (Grzymala 1994). All these techniques, however, try to solve this case in an automatic way without taking into account possible information coming from the decision maker.

The belief that the decision maker may have his own *preferences* to acceptable and non-acceptable matching between the rule and the object is the basis of our system. We are going to incorporate him in the process of determining the classification suggestions. It is done by using the idea of the so called *valued closeness relation* introduced in (Slowinski 1993) and (Slowinski and Stefanowski1993).

The valued closeness relation is constructed by using similar principle as the valued outranking relation introduced in (Roy 1985). According to this principle, the new object is compared to each decision rule in order to assess the credibility of the affirmation : "a rule is close to an object". The formula for the calculation of the credibility is essentially based on two tests: *concordance* and *discordance*. While constructing the both tests the decision maker may express his preferences. He is able to express a *relative importance* that he wishes to assign to each attribute. Moreover, for each attribute (condition) he can define, using special thresholds, his meaning of possible *indiscernibility, small difference* or *strong difference* between an object and a rule. Due to a limited size of the paper, it is not possible to present the details of formulae for calculations of the closeness degree. All necessary information can be found in (Slowinski and Stefanowski 1993).

In classification support all rules are scanned and for each of them the degrees of closeness are calculated. Then, the rules with the greatest values of the degrees are presented to the decision maker together with information about the strength of corresponding decisions. The decision maker must interpret these information. As a default suggestion, the strongest decision is taken into account.

Table 1: Performance (accuracy of classification in %) of VCR-Class

data set	VCR–Class		
	phase 3	phase 2	phase 1
large soybean	87.9	85.7	79.2
election	89.4	79.5	71.8
iris	95.3	89.3	88
hsv4	58.2	49.2	41.9
hsv2	77.1	70.5	59.8
concretes	88.9	82.8	81
breast cancer	67.1	59.3	51.2
imidasolium	53.3	44.8	34.4
lymphograpy	85.2	73.6	67.6
oncology	83.8	82.4	74.1
buses	98	93.5	90.8
bearings	96.4	90.9	87.3
small soybean	97.8	97.8	97.8

3 Experiments

The presented idea of using valued closeness relation in classification support has been implemented as an interactive software system called *VCR - Class.* This program uses the set of rules induced from examples by means of the modified *LEM* procedure available in the program *RoughDAS* (Slowinski and Stefanowski 1992).

Several computation experiments were performed to evaluate the *VCR - Class* system. In these experiments, the correctness of classification was verified. It was done on the basis of the *accuracy of classification* (expressed in percentage) obtained in "*10-fold-cross-validation*" or "*leave-one-out*" tests. Possible preferences of the decision maker were simulated and classification tests were performed in an automatic way.

Several different real-life data sets were used for the experiments. They were coming from known applications of rough set theory or they were well-known sets of examples from machine learning literature. The data sets were the following: *large and small soybean disease, iris, breast cancer, lymphography, election, highly selective vagotomy* (two versions: *hsv4, hsv2*), *imidasolium, buses, rolling bearings, frost resistant concretes* (denoted further as *concretes*) and *oncology.* First four data sets were obtained from the *UCI repository of machine learning databases* at the University of California at Irvine (Murphy and Aha 1991). The author is grateful to the creators of these databases and all other people who allowed him to use their data sets for these experiments.

Let us notice that the data sets used in experiments were completely defined, i.e. they did not contain missing values. Due to this fact some of the data were modified by removing certain examples or attributes.

As the main aim of the experiment was to check the usefulness of the proposed strategies of decision

Table 2: Performance (accuracy of classification in %) of compared systems; — denotes impossibility of performing a classification test for PRISM because of inconsistent examples; ? denotes impossibility of performing "leave-one-out" test for a given implementation.

data set	compared systems					
	ID3	PT	PRI-SM	ELY-SEE	C4.5 rules	VCR -Class
large soybean	81.2	76.6	62.5	86.8	88.4	87.9
election	84.4	88.8	76.7	84.0	89.6	89.4
iris	90.7	90.7	90	94.7	91.3	95.3
hsv4	50	60.7	—	52.5	61.4	58.2
hsv2	68	71.3	—	68	78.1	77.1
concret.	86.6	92	—	89.4	87.1	88.9
breast cancer	62.5	68.1	—	63.5	68.1	67.1
imidas.	35.3	35.3	35.8	59.7	52.1	53.3
lymph.	75	82.4	66.9	79.7	80.4	85.2
oncol.	78.6	84	73.2	81.9	79.8	83.8
buses	94.7	96.9	?	?	97.4	98
bearings	83.6	85.4	?	?	83.8	96.4
small soybean	97.8	97.8	?	?	97.8	97.8

supporting, the classification of testing examples was done in three phases. In the first phase, classification suggestions were determined on the basis of exact matching. In the second phase, the strategies of solving multiple ambiguous matching were also used. In the third phase, beside the two above techniques, the strategy with valued closeness relation was additionally taken into account. Results are presented in Table 1.

Moreover, in the experiments we compared the accuracy of classification obtained using *VCR–Class* system with other systems known from machine learning field. Thus, for the same data sets (in fact, for the same division of them into training and testing parts) the classification tests were performed by using implementations of: *PRISM* algorithm (Cendrowska 1987), classical version of Quinlan's *ID3*, *PT* – its modification with pre-pruning (similar idea as in (Cestnik *et al* 1987), *ELYSEE* method (Teghem 1992), and *C*4.5 system in a rule option (Quinlan 1993). The results are presented in Table 2.

The above results show that the proposed strategies (both handling ambiguity in matching and looking for nearest rules) are very effective. Using them in all experiments significantly increased the accuracy of classification (in many data sets about 20%; see Table 1). On the other hand, one can also claim that restricting the rule classification system to simple matching (i.e. phase 1) leads to its unsatisfactory performance (see also results for *PRISM* in Table 2). Similar conclusions for different techniques have been also obtained

Table 3: Performance of VCR-Class systems with reduced and non-reduced sets of attributes; acc. means accuracy of classification in %, attr. - number of attributes used in induction, rul. - number of rules in the obtained set of rules

data set	all attributes			reduced attributes		
	acc.	attr.	rul.	acc.	attr.	rul.
election	89.4	30	48	89.0	7	93
hsv4	58.2	11	53	57.3	5	47
hsv2	77.1	11	32	76.2	5	37
concretes	88.9	10	35	85.0	6	41
breast can.	67.1	8	109	66.0	7	116
imidas.	53.3	8	125	52.1	4	111
oncology	83.8	17	40	80	15	46
buses	96.1	8	6	98.0	3	5
bearings	83.6	10	35	96.4	6	41

in (Grzymala 1994). One can also notice that accuracy obtained by the *VCR–Class* system is comparable with the results obtained by other well-known systems.

Moreover, for some of the studied data sets we additionally checked the possibility of *attribute reduction* in the input data set. It was done on the basis of reducts found using rough set approach (implemented in *RoughDAS* system. It must be stressed, however, that the concept of the reduct is defined for the closed world assumption (Pawlak 1991), i.e. without taking into account possible new objects. Thus, one should not expect that such a reduction can automatically lead to the better performance of *VCR–Class* system. The results presented in Table 3 confirmed this observation. Only for two data sets the prior reduction increased the accuracy while for other decreased it.

4 Conclusions

The main objective of this paper was to evaluate the usefulness of the valued closeness concept in computation experiments on real-life data sets. The obtained results show that such concepts are necessary for rule driven classification systems.

5 Acknowledgements

The author is very grateful to Professors Roman Slowinski and Jerzy Grzymala-Busse for discussions on the problems considered in this paper. The work on this paper has been supported by KBN grant no. 8 S503 016 06.

References

Cendrowska J. 1987. "PRISM, an algorithm for inducing modular rules." *Int. J. Man-Machine Studies*, **27**, 349-370.

Cestnik B., Kononenko I., Bratko I. 1987." Assistant 86, a knowledge elicitation tool for sophisticated users." In Bratko I., Lavrac N. (eds) *Progress in Machine Learning*, Sigma Press, Wilmshow, 31-45.

Clark P., Niblett T. 1989. "The CN2 induction algorithm." *Machine Learning*, **3**, 261-283.

Grzymala-Busse J.W. 1992. "LERS - A system for learning from examples based on rough sets." In Slowinski R. (ed.): *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory*, Kluwer Acad. Publ., 3-18.

Grzymala-Busse J.W. 1994. "Managing uncertainty in machine learning from examples." *Proc. of 3rd Int. Symp. on Intelligent Systems*, Wigry Poland, June 1994, (Springer Verlag, in press).

Murphy P.M., Aha, D.W. 1991. UCI Repository of machine learning databases. [Machine-readable data repository]. Irvine CA. University of California, Dept. of Computer Sci.

Michalski R.S., Mozetic I., Hong J., Lavrac N. 1986. "The multi-purpose incremental learning system and its testing application to three medical domains." *Proc. of 5th Nat. Conf. on AI*, 1041-1045.

Pawlak Z. 1991. *Rough Sets. Theoretical Aspects of Reasoning about Data.* Kluwer Academic Publ.

Quinlan J.R. 1993. *C4.5: Programs for Machine Learning.* Morgan Kaufmann CA.

Roy B. 1985. *Methodologies Multicritere d'Aide a la Decision.* Economica, Paris.

Slowinski R. 1993. "Rough sets learning of preferential attitude in multi-criteria decision making." In Komorowski J., Ras Z.W. (eds), *Proc. of Int. Symp. on Methodologies for Intelligent Systems*, Springer Verlag LNAI 689, 642-651.

Slowinski R., Stefanowski J. 1992. " 'RoughDAS' and 'RoughClass' software implementation of the rough set approach." In Slowinski R. (ed.): *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory*, Kluwer Acad. Publ. 445-456.

Slowinski R., Stefanowski J. 1993. "Rough Classification with Valued Closeness Relation." In Didey E. et al.(eds) : *New Approaches in Classification and Data Analysis* Springer Verlag, Studies in Classification, Data Analysis and Knowledge Organization, 482-489.

Stefanowski J. 1993. "Classification and Decision Supporting Based on the Rough Set Theory." *Foundations of Computing and Decision Sciences*, **18**, 371-380.

Teghem J., Benjelloun M. 1992. "Some experiments to compare rough sets theory and statistical methods." In Slowinski R. (ed.): *Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory*, Kluwer Acad. Publ., 165-180.

Weiss S. M., Kulikowski C.A. 1991. *Computer Systems That Learn: Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning and Expert Systems*, Morgan Kaufmann.