classification, decision tables,
nondeterministic decision rules,
rough sets, rule-based classifiers

Piotr PASZEK[1], Barbara MARSZAŁ-PASZEK[1]

DETERMINISTIC AND NONDETERMINISTIC DECISION RULES IN CLASSIFICATION PROCESS

In this paper an algorithm of calculating nondeterministic decision rules from the decision table was presented. The algorithm uses additional conditions imposed on rules. This is a greedy algorithm. The nondeterministic decision rules were used in the process of classification of new examples, for medical data sets. The decision tables from the UCI Machine Learning Repository were used. The achieved results allow us to state that nondeterministic decision rules can be used for improving the quality of classification.

1. INTRODUCTION

Over the years many methods based on rule induction and rule-based classification systems were developed [14,20]. The part of these systems found application in diagnosis support systems, medical expert systems and object classification [5,6,8,10,11,14,16,23,27,28,29]. Some of them are based on rough sets [4,9,17,21,22,26]. In this paper we show that there is still room for improving the rule-based classification systems. We discuss a method for rule inducing based on searching for strong rules for a union of a few relevant decision classes - nondeterministic decision rules.

In the paper, the following classification problem is considered: for a given decision table *T* [18,19] and a new object *v* generate a value of the decision attribute on *v* using values of conditional attributes on *v*.

In [24,25] Skowron and Suraj shown that there exist information systems $S=(U,A)$ [18], where U is a finite set of objects and A is a finite set of attributes, such that the set U can't be described by deterministic rules. In [15] Moshkov shown that for any information system, the set U can be described by nondeterministic (inhibitory) rules. Inhibitory rules [7] are a special case of nondeterministic rules. These results inspired us to use the nondeterministic rules [13] in the classification process.

We present an application of (bounded) nondeterministic rules in construction of rule-based classifiers. We include the results of experiments showing that by combining rule-based classifiers based on minimal decision rules [12,19] with the nondeterministic rules having the sufficiently large support [1] it is possible to improve the classification quality and reduce the classification error. Experiments were done on decision table from medical domain. Reducing the classification error is significant especially in diagnosis support systems. In such systems every classification error is connected with consequences for the patient.

The paper consists of six sections. In Section 2, we recall the notions of a decision table and deterministic and nondeterministic decision rules. In Sections 3 and 4 we present a greedy algorithm for nondeterministic decision rule construction and main steps in construction of classifiers enhanced by nondeterministic rules. In Section 5 results of experiments with real-life data from medical domain are discussed. Section 6 contains short conclusions.

[1] Institute of Computer Science, University of Silesia, Będzińska 39, 41-200 Sosnowiec, Poland
e-mail: paszek@us.edu.pl

2. MAIN NOTIONS

In this section main notions for nondeterministic decision rules are described.

2.1. DECISION TABLES

Let $T = (U, A, d)$ be a *decision table* [18], where $U = \{u_1, \ldots, u_n\}$ is a finite nonempty set of *objects*, $A = \{a_1, \ldots, a_m\}$ is a finite nonempty set of *conditional attributes* and d is the *decision attribute*. We assume that for each $u_i \in U$ and each $a_j \in A$ the value $a_j(u_i)$ and the value $d(u_i)$ belong to ω, where $\omega = \{0,1,2,\ldots\}$ is the set of nonnegative integers. By $V_d(T)$ we denote the set of values of the decision attribute d on objects from U.

2.2. DETERMINISTIC DECISION RULES

Let us consider a rule

$$(a_{j_1}(x) = b_1) \wedge \ldots \wedge (a_{j_t}(x) = b_t) \Rightarrow (d(x) = b),$$

where $a_{j_1}, \ldots, a_{j_t} \in A$, $b_1, , \ldots, b_t \in \omega$, $b \in V_d(T)$ and numbers $j_1, \ldots, j_t$ are pairwise different. Such rules are called *deterministic decision rules*.

2.3. NONDETERMINISTIC DECISION RULES

In general, nondeterministic decision rules in a given decision table T are of the form

$$(a_{j_1}(x) = b_1) \wedge \ldots \wedge (a_{j_t}(x) = b_t) \Rightarrow (d(x) = c_1) \vee \ldots \vee (d(x) = c_s), \quad (1)$$

where $a_{j_1}, \ldots, a_{j_t} \in A$, $b_1, \ldots, b_t \in \omega$, numbers $j_1, \ldots, j_t$ are pairwise different, and $\emptyset \neq \{c_1, \ldots, c_s\} \subseteq V_d(T)$. We consider nondeterministic rules with cardinality $|\{c_1, \ldots, c_s\}|$ small in comparison with $|V_d(T)|$.

Let us introduce some notation. If r is the nondeterministic rule of the form (1) then by α we denote its left hand side, i.e., the formula $(a_{j_1} = b_1) \wedge \ldots \wedge (a_{j_t} = b_t)$, and by β its right hand side, i.e., the formula $(d = c_1) \vee \ldots \vee (d = c_s)$. By $\|\alpha\|_T$ (or $\|\alpha\|$, for short) we denote all objects from U satisfying α.

To measure the quality of such rules we use coefficients called the *support* and the *confidence* [1]. They are defined as follows. If r is a nondeterministic rule of the form (1) then the support of this rule in the decision table T is defined by

$$supp(r) = \frac{\left\| \|\alpha\| \cap \|\beta\| \right\|}{|U|},$$

and the confidence of r in T is defined by

$$conf(r) = \frac{\left\| \|\alpha\| \cap \|\beta\| \right\|}{\|\alpha\|}.$$

We also use a normalized support of r in T defined by

$$n\_supp(r) = \frac{supp(r)}{\sqrt{|\{c_1, \ldots, c_s\}|}}.$$

Now we can define a set of nondeterministic decision rules which are used in Section 4 for enhancing the quality of classification of rule-based classifiers. This set is defined relative to the following three parameters:

1. $\alpha \in (0.5,1]$ - a threshold used as the lower bound for the confidence of rules;
2. $n\_sup \in (0,1]$ - a threshold used as the lower bound for the normalized support of rules;
3. k - a threshold used as an upper bound on the number of decision values on the right hand sides of rules; in our heuristic method k is assumed to be small.

The set of nondeterministic rules $Rule_{nd}(\alpha, n\_sup, k)$ is defined as the subset of all nondeterministic rules r (over attributes in T) such that

1. $conf_T(r) \geq \alpha$;
2. $norm\_supp_T(r) \geq n\_sup$ and;
3. $|V(r)| \leq k$.

The algorithm presented in Section 3 is searching for nondeterministic rules with sufficiently large support and relatively small (in comparison to the set of all possible decisions), the sets of decisions defined by the right hand sides of such rules for the decision table *T*. Such rules are combined with minimal rules [12,19] for increasing the classification quality. The details are presented in Section 5.

3. ALGORITHM FOR NONDETERMINISTIC DECISION RULE CONSTRUCTION

Let us describe the algorithm with threshold $\alpha \in (0.5,1]$ and k which constructs the nondeterministic decision rules for T. This algorithm is based on greedy strategy which is used to minimize the length of rules.

First, the minimal rules are constructed for a given decision table T.

Next, these rules are shortened.

Greedy algorithm for nondeterministic decision rule construction Rul_{nd}

Input: decision table *T*, real number $0.5 < \alpha \leq 1$ and threshold k - upper bound on the number of decision values.

Output: $Rule_{nd}(\alpha, n\_sup, k)$ a set of nondeterministic decision rules for T.

begin
    Generate *Rul* a set of minimal decision rules of T
    $Rul_{nd} \leftarrow \varnothing$
    for $R \in Rul$ **do** $\{R: T \rightarrow (d = v), T = D_1 \wedge D_2 \wedge \cdots \wedge D_m, v \in V_d\}$
        Stop $\leftarrow$ *false;*
        $\lambda_R \leftarrow supp(R)$;
        repeat
            for $D_i \in T$ **do**
                $r_i \leftarrow r$ $\{r_i$ is obtained by dropping D_i from the left hand side of rule $r\}$
                $\|T^i\|_T$;
                $\vartheta = \{v \in V_d : \exists_{x \in U_{Ti}} d(x) = v\}$;
                Choice $\theta_i \{\theta_i \subseteq \vartheta : conf(T^i \rightarrow (d = \theta_i))\} \geq \alpha$; {greedy}
                $\lambda_{T^i} \leftarrow supp(T^i \rightarrow \theta_i)/\sqrt{|\theta_i|}$;
            endfor
            $\lambda_{max} \leftarrow argmax\{\lambda_{T^i}\}$;
            if $\lambda_{max} \geq \lambda_R$
                then $R \leftarrow R_s \{R_s : T \rightarrow (d = \theta_i), \lambda_T\}$
                else Stop $\leftarrow$ *true*
            endif
        until Stop
        if $\|\theta_i\| \leq k$
            then $Rul_{nd} \leftarrow Rul_{nd} \cup \{R_s\}$;
    endfor
end

4. CLASSIFIERS BASED ON NONDETERMINISTIC DECISION RULES

In this section, we present an application of nondeterministic rules for classification of objects.

The set of nondeterministic rules and the set of minimal rules generated by the system RSES [4] build our classifier. Because we have two groups of rules in the classification process we should negotiate between them. For any new object the decision value set is generated as follows.

First, for any new object, all nondeterministic rules matching the object are extracted. Next, from these matched rules, a rule with the largest (normalized) support is selected. In the case when several rules have the same support, the decision value set $V(r)$ of the nondeterministic rule r with the smallest set of decision value is selected. If still several nondeterministic rules with the above property exist then one of them is selected randomly.

Next, for this object, all minimal rules matching the object are extracted. We obtain a single decision value c using standard voting procedure.

In this way, for any new object we obtain a decision value c and a decision value set $V(r)$, where r is the rule selected from the set of nondeterministic rules.

The final decision for a given new object is obtained from the decision c and decision value set $V(r)$. This decision is defined by the following strategy to resolve conflicts [13].

1. If for a given new object the standard voting based on minimal rules predicts the decision value c and $c \in V(r)$, (i.e., no conflict arises) then as the final decision the single decision c we take.
2. If for a given new object the standard voting based on minimal rules predicts the decision value c and $c \notin V(r)$ (i.e., conflict arises) then we take as the final decision value the single decision value c provided the minimal rule supports larger than the normalized support of the decision rule r generated by the algorithm and selected for the given new object. In the opposite case, we take as the final decision a single decision value from the set $V(r)$, with the largest support in T among decisions from $V(r)$.
3. If for a new object, the standard voting based on minimal rules predicts the decision value c and this object does not match any rule generated by the algorithm then we assign the decision c as the final decision.
4. If a given new object does not match any of the minimal rules then we assign as the final decision the single decision from $V(r)$ with the largest support among decisions from $V(r)$, where r is the rule selected by voting on nondeterministic rules.
5. In the remaining cases, a given new object is not classified.

5. EXPERIMENTS

We have performed experiments on decision tables from [3] using classification algorithms C. The classification algorithm C is obtained by the described above combination of the auxiliary classification algorithm from RSES based on all minimal decision rules with the classification algorithm based on nondeterministic rules, described in previous section.

The majority of decision tables used for experiments concern medical data.

Decision table *Dermatology* contains data about the diagnosis of erythemato-squamous diseases, a real problem in dermatology [8]. Decision table *Ecoli* concerns the protein localization sites in Escherichia coli bacteria [16]. The classification task of decision table *Postoperative* is to determine when patients in a postoperative recovery area should be sent to the next one [1].

Lymphography and *Primary Tumor* data are two of three domains provided by the University Medical Centre, Institute of Oncology from Ljubljana that has repeatedly appeared in the machine learning literature [6,14].

Some attributes in decision tables used for experiments were discretized, and missing values were filled by algorithms from RSES. In evaluation of the accuracy of classification algorithms on a decision table (i.e., the percentage of correctly classified objects) the cross-validation method was used.

For any considered data table, we used the classification algorithms *C* for different values of parameter α. On testing sets the accuracy and the coverage factor were calculated. Also the *maximal relative deviation* (mrd) was calculated.

Table 1 contains the results of our experiments. For all (seven) decision tables the classification quality measured by *accuracy* × *coverage* was better for the classification algorithm *C* than in the case of the classification algorithm from RSES based only on minimal rules with standard voting.

For four decision tables, the *mrd* was no greater than 5% in the case when we used the classification algorithm *C*. Hence, using the classification algorithm *C* may lead to more stable classification.

Table 1. Results of experiments with deterministic and nondeterministic rules

Decision Table	Classification Factor	Classification algorithm					
		Alg(1)	*C, α*(2)				
			1.0	0.9	0.8	0.7	0.6
Dermatology	acc × cover	95.17	**95.26**	91.35	87.07	86.61	82.88
	mrd	0.036	0.035	0.026	0.018	0.025	0.054
Ecoli	acc × cover	53.35	59.45	**60.61**	60.27	60.42	56.25
	mrd	0.043	0.026	0.031	0.047	0.049	0.037
Iris	acc × cover	61,31	73,67	**74,89**	73,87	**74,89**	74,78
	mrd	0,073	0,070	0,069	0,052	0,042	0,059
Lymphography	acc × cover	12,30	27,40	28,07	**29,43**	28,28	29,29
	mrd	0,053	0,077	0,104	0,091	0,109	0,092
Postoperative	acc × cover	16,81	**69,19**	68,19	67,44	66,26	65,00
	mrd	0,037	0,036	0,082	0,107	0,294	0,294
Primary Tumor	acc × cover	65.29	65.49	**66.08**	66.08	66.08	
	mrd	0.188	0.185	0.174	0.174	0.174	
Zoo	acc × cover	89.87	90.07	**90.63**	90.50	80.63	80.66
	mrd	0.037	0.059	0.074	0.043	0.055	0.055

(1) In the column marked by *Alg* the classification is defined by the classification algorithm from RSES based on all minimal rules.

(2) Confidence of nondeterministic rules generated by the algorithm is not smaller than the parameter α.

For obtaining those better results, it was necessary to optimize the threshold α for each data table. This means that the parameter α should be tuned to the data table.

6. CONCLUSIONS

Results of experiments show that the classification algorithms based on nondeterministic rules is better than that based on deterministic decision rules. This means that nondeterministic decision rules are as relevant to classification algorithms as deterministic decision rules.

There is an additional motivation for the use of nondeterministic decision rules in classification algorithms: the nondeterministic decision rules have much more chance to have larger support than the deterministic ones. Therefore they are more often accepted by experts, particularly in medical expert systems or diagnosis support systems.

Using nondeterministic rules in a decision support system can lead to improving the classification quality, and to reducing the terror rate. This is very important especially in diagnosis support systems. In such systems every classification error can be connected with serious consequences for the patient.

BIBLIOGRAPHY

[1] BUDIHARDJO A., GRZYMALA–BUSSE J., WOOLERY L., Program LERS_LB 2.5 as a tool for knowledge acquisition in nursing, In: Proceedings of the 4th Int. Conference on Industrial & Engineering Applications of AI & Expert Systems, 1991, pp. 735–74.

[2] AGRAWAL R., IMIELINSKI T., SWAMI A., Mining Association Rules Between Sets of Items in Large Databases. In: P. Buneman, S. Jajodia (eds.), Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington, D.C., New York, ACM Press, 1993, pp. 207–216.

[3] ASUNCION A., NEWMAN D.J., UCI Machine Learning Repository. University of California, Irvine, School of Information and Computer Sciences, 2007.

[4] BAZAN J.G., SZCZUKA M.S., WOJNA A., WOJNARSKI M., On the Evolution of Rough Set Exploration System. In: S. Tsumoto et. al. (eds.), RSCTC 2004, LNAI, Vol. 3066, Springer, Heidelberg, 2004, pp. 592–601.

[5] CARLIN U., KOMOROWSKI J., OHRN A., Rough set analysis of patients with suspected acute appendicitis. In: B. Bouchon–Meunier, R.R. Yager, (eds.), Proc. Seventh Conference on Information Processing and Management of Uncertainty in Knowledge–Based Systems, (IPMU'98). Paris, France, Edisions, 1998, pp. 1528–1533.

[6] CESTNIK G., KONENENKO I., BRATKO I., Assistant–86: A Knowledge–Elicitation Tool for Sophisticated Users. In: I.Bratko, N. Lavrac (eds.) Progress in Machine Learning, Sigma Press, 1987, pp. 31–45.

[7] DELIMATA P., MOSHKOV M., SKOWRON A., SURAJ Z., Inhibitory Rules in Data Analysis: A Rough Set Approach. Studies in Computational Intelligence, Vol. 163, Springer, Heidelberg, 2009.

[8] DEMIROZ G., GOVENIR H. A., ILTER N., Learning Differential Diagnosis of Eryhemato–Squamous Diseases using Voting Feature Intervals, Aritificial Intelligence in Medicine, Vol. 13, 1998, pp. 147.

[9] GRZYMALA–BUSSE J.W., LERS – A Data Mining System. In: Maimon, O., Rokach, L. (eds.), The Data Mining and Knowledge Discovery Handbook. Springer, New York, 2005, pp. 1347–1351.

[10] GRZYMALA–BUSSE J.W., GOODWIN L.K., A comparison of less specific versus more specific rules for preterm birth prediction, Proceedings of the First Online Workshop on Soft Computing WSC1 on the Internet, Japan, 1996, pp. 129–133.

[11] JELONEK J., KRAWIEC K., SLOWINSKI R., STEFANOWSKI J., SZYMAS J., Neural networks rough sets – Comparison combination for classification of histological pictures, In: W. Ziarko (ed.), Rough Sets, Fuzzy Sets and Knowledge Discovery (RSKD'93), Springer–Verlag & British Computer Society, London, Berlin, 1994, pp. 426–433.

[12] KOMOROWSKI J., PAWLAK Z., POLKOWSKI L., SKOWRON A., Rough sets: A tutorial, in S.K. Pal and A. Skowron (eds.), Rough–Fuzzy Hybridization: A New Trend in Decision–Making, Springer–Verlag, Singapor, 1999, pp. 3–98.

[13] MARSZAŁ–PASZEK B., PASZEK P., Minimal Templates and Knowledge Discovery. In: M. Kryszkiewicz et al. (eds.), RSEISP 2007. LNAI, Vol. 4585, Springer, Heidelberg, 2007, pp. 411–416.

[14] MICHALSKI R., MOZETIC I., HONG J., LAVRAC N., The Multi–Purpose Incremental Learning System AQ15 and its Testing Applications to Three Medical Domains. Proceedings of the Fifth National Conference on Artificial Intelligence,.Philadelphia, Morgan Kaufmann, 1986, pp. 1041–1045.

[15] MOSHKOV M., SKOWRON A., SURAJ Z., Maximal consistent extensions of information systems relative to their theories. Information Sciences 178 (12), 2008, pp. 2600–2620.

[16] NAKAI K., KANEHISA M., Expert Sytem for Predicting Protein Localization Sites in Gram–Negative Bacteria, Proteins: Structure, Function, and Genetics 11, 1991, pp. 95–110.

[17] NGUYEN S.H., SKOWRON A., SYNAK P., WRÓBLEWSKI J., Knowledge Discovery in Databases: Rough Set Approach, Seventh International Fuzzy Systems Association World Congress (IFSA'1997), Vol. 2, Academia, Prague, 1997, pp. 204–209.

[18] PAWLAK Z., Rough Sets: Theoretical aspects of reasoning about data. Boston: Kluwer Academic Publishers, 1991.

[19] PAWLAK Z., SKOWRON A., Rudiments of Rough Sets. Information Sciences 177, 2007, pp. 3–27; Rough Sets: Some Extensions. Information Sciences 177, pp. 28–40; Rough Sets and Boolean Reasoning. Information Sciences 177, pp. 41–73.

[20] PULATOVA S., Covering (Rule–Based) Algorithms. In: Berry, M.W., Browne, M. (eds.), Lecture Notes in Data Mining. World Scientific, Singapore, 2006, pp. 87–97.

[21] *Rosetta*: http://www.lcb.uu.se/tools/rosetta/

[22] *Rough Set Exploration System*: http://logic.mimuw.edu.pl/~rses.

[23] SWINIARSKI R., Rough sets Bayesian methods applied to cancer detection, In: L. Polkowski, A. Skowron (eds.), Proc. First International Conference on Rough Sets and Soft Computing RSCTC'98. Warszawa, Poland, LNAI 1424, Springer–Verlag, 1998, pp. 609–616.

[24] SKOWRON A., SURAJ Z., Rough sets and concurrency, Bulletin of the Polish Academy of Sciences 41 No. 3, 1993, pp. 237–254.

[25] SURAJ Z., Some Remarks on Extensions and Restrictions of Information Systems, In: W. Ziarko, Y.Y. Yao (eds.), RSCTC 2000. LNCS, Vol. 2005. Springer, Heidelberg, 2001, pp. 204–211.

[26] TRIANTAPHYLLOU E., FELICI G. (eds.), Data Mining and Knowledge Discovery Approaches Based on Rule Induction Techniques, Springer Science and Business Media, LLC, New York, 2006.

[27] TSUMOTO S., Domain experts' interpretation of rules induced from clinical databases. In: H.J. Zimmermann (ed.), Proceedings of the Fifth European Congress on Intelligent Techniques and Soft Computing (EUFIT'97), Aachen, Germany, Verlag Mainz, 1997, pp. 1639–1642.

[28] WAKULICZ–DEJA A., PASZEK P., Diagnose Progressive Encephalopathy Applying the Rough Set Theory, International Journal of Medical Informatics, 46, 1997, pp. 119–127.

[29] WOOLERY K., GRZYMAŁA–BUSSE J., Machine learning for an Expert System to Predict Preterm Birth Risk, J Am Med Informatics Assoc 1, 1994, pp. 439–446.