ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 206, 2007

*Iwona Gruszka**

THE ROUGH SETS APPROACH TO MULTICRITERIA EU'S COUNTRIES CLASSIFICATION PROBLEM BASED ON DOMINANCE RELATION – THE PROBABILISTIC CHARACTERISTICS OF DECISION RULES

Abstract. The rough sets theory was introduced by Z. Pawlak (1982). The mathematical base on rough sets approach is a binary relation on universe of objects. In the classic rough sets theory there is an indiscernibility relation. As an equivalence relation it permit to divide the universe of objects on equivalence classes called elementary sets and forms a basic granules of knowledge of the universe. For creating good decision model (with possible small number of robust rules) the granulation process is indispensable. However, from the other point of view, it is natural to extend the indiscernibility concept taking into account the situations where some objects dominate another ones by the considered criteria which domains are preferentially ordered.

For this reason S. Greco, B. Materazzo and R. Słowiński (1996a, b, 1999) have proposed an extension of the rough set theory. This innovation is based on substitution of the indiscernibility relation by a dominance relation in the rough approximation of decision classes.

The aim of this work is the decision analysis of EU's countries classification problem for designing the decision model with dominance relation approach using the "4eMka" system. Also the probabilistic characteristics of decision rules are presented.

Key words: rough sets, multicriteria classification problem.

1. INTRODUCTION

Rough sets theory was introduced by Z. Pawlak (1982). The rough set philosophy is founded on the assumption that with every objects of the universe of discourse we associate some information (knowledge). Objects characterized by the some information are indiscernible in view of the

* M.Sc., Karol Adamiecki University of Economics, Katowice.

available information about them. The indiscernibility relation generated in this way is the mathematical basis for the rough sets theory.

Any set of all indiscernible objects is called elementary set, and forms a basic granule of knowledge about the universe. Any set of objects being a union of some elementary sets is referred to as crisp (precise), otherwise a set is rough (imprecise). Consequently, each rough set has boundary zone, i.e. objects which cannot be classified with certainty as members of the set or of its complement. Therefore the rough set can be replaced by a pair of crisp sets, called the lower and upper approximation. The lower approximation consists of all objects which surely belong to the set and the upper approximation contain objects which possibly belong to the set.

Classic definition of lower and upper approximations (Pawlak 1982, 1999) were originally introduced with reference to an indiscernibility relation which was assumed to be an equivalence relation (reflexive, symmetric and transitive). It is quiet natural to extend the indiscernibility concept to take account of situations where ones objects dominates another ones by the considered criteria which domains are preferentially ordered.

For this reason S. Greco, B. Materazzo and R. Słowiński (1996a, b, 1999) have proposed an extension of the rough set theory. This innovation is based on substitution of the indiscernibility relation by a dominance relation in the rough approximation of decision classes. Such relation express weaker form of indiscernibility and, usually, are not equivalence relation. While the reflexivity and transitivity property seems quite necessary to express any form of dominance, the symmetry properties may be relaxed.

Also, by Z. Pawlak's (2001) research, a probabilistic support for decision rules are possible.

This paper is organized as follows. In the first step the idea of dominance relation is presented and its use to rough sets analysis of classification problem. Also, in the first chapter, the probabilistic characteristics are defined. The second part contains the results of rough sets approach based on dominance relation to EU's countries classification problem. The final section groups conclusions.

2. THE ROUGHT SETS APPROACH BASED ON DOMINANCE RELATION TO CLASSIFICATION PROBLEM

Rough set based data analysis starts from a data table, called an decision table. The decision table contains data about objects of interests evaluated in terms of some criteria. We distinguish in the decision table condition (C) and decision (d) criteria. The condition criteria provide as the information about the universe of considered objects, while the decision criteria describes the classification decision made by decision maker (an *a priori* classification). The decision table describes decision in terms of conditions that must be satisfied in order to carry out the specified decisions.

With every decision table we can associate a set of decision rules, which forms a decision model and which can be seen as a logical description of approximations of decisions. Also, each of the decision rule can be described by some probabilistic properties. The decision situation can be placed in a probabilistic space, then, and these properties give a new look of drawing conclusion from data.

Let $S = \langle U, C \cup \{d\}, V, f \rangle$ is a decision table,
where:

U – a finite set of objects (universe),
C – a finite set of condition criteria,
d – a decision criterion,
V – a criterion's domain: $V = \bigcup_{c \in C} V_c$,
V_c – a domain of criterion c,
f – an information function: $f: U \times Q \to V$ such that $f(x, c) \in V_c$ for each $c \in C$, $x \in U$.

Let S_c be an outranking relation on U with reference to criterion $c \in C$, such that xS_cy means "object x is at least as good as y with respect to criterion c". Suppose that S_c is a complete preorder, that is reflexive and transitive binary relation. Moreover, let $Y = \{Y_t : t \in T\}$, $T = 1, ..., n$, be a set of classes of U, such that each $x \in U$ belongs to one and only one of class $Y_t \in Y$. We assume that for every $r, s \in T$, such that $r > s$, each object of Y_r is preferred to each object of Y_s. More formally, if S is a comprehensive outranking relation on U, i.e. xSy means: "object x is at least as good as y", $x, y \in U$, then it is supposed that $[x \in Y_r, y \in Y_s, r > s] \Rightarrow [xSy, \neg ySx]$.

Let us also consider the following upward and downward cumulated classes, respectively:

$$Y_t^{\leqslant} = \bigcup_{s \leqslant t} Y_s \tag{1}$$

$$Y_t^{\geqslant} = \bigcup_{s \geqslant t} Y_s \tag{2}$$

Observe also that $Y_1^{\geqslant} = Y_n^{\leqslant} = U$, $Y_n^{\geqslant} = Y_n$ and $Y_1^{\leqslant} = Y_1$.

It is said that x dominates y with respect to C (denotation xD_cy), if $xS_cy, c \in C$.

Given $x \in U$, let us describe the C-dominating set of x and C-dominated set of x respectively

$$D_C^+(x) = \{y \in U : yD_C x\} \tag{3}$$

$$D_C^-(x) = \{y \in U : xD_C y\} \tag{4}$$

Note, that the sets D_C^+ and $D_C^-(x)$ forms a basic granules of knowledge about the universe of analyzed objects.

We can define the C-lower and the C-upper approximation of $Y_t^{\geqslant}$ (denotation $\underline{C}Y_t^{\geqslant}$) and $Y_t^{\geqslant}$ (denotation $\overline{C}Y_t^{\geqslant}$), $t \in T$, respectively as

$$\underline{C}Y_t^{\geqslant} = \{x \in U : D_C^+(x) \subseteq Y_t^{\geqslant}\} \tag{5}$$

$$\overline{C}Y_t^{\geqslant} = \bigcup_{x \in Y_t^{\geqslant}} D_C^+(x) \tag{6}$$

Analogously, we define the C-lower and the C-upper approximation of $Y_t^{\leqslant}$ (denotation $\underline{C}Y_t^{\leqslant}$) and $Y_t^{\leqslant}$ (denotation $\overline{C}Y_t^{\leqslant}$), $t \in T$, respectively as

$$\underline{C}Y_t^{\leqslant} = \{x \in U : D_C^-(x) \subseteq Y_t^{\leqslant}\} \tag{7}$$

$$\overline{C}Y_t^{\leqslant} = \bigcup_{x \in Y_t^{\leqslant}} D_C^-(x) \tag{8}$$

The C-lower and the C-upper satisfy the following properties, $t \in T$:

$$\underline{C}Y_t^{\geqslant} \subseteq Y_t^{\geqslant} \subseteq \overline{C}Y_t^{\geqslant} \tag{9}$$

$$\underline{C}Y_t^{\leqslant} \subseteq Y_t^{\leqslant} \subseteq \overline{C}Y_t^{\leqslant} \tag{10}$$

The C-boundaries (C-doubtful regions) of $Y_t^{\geqslant}$ and $Y_t^{\leqslant}$, $t \in T$ are defined as

$$BN_C(Y_t^{\geqslant}) = \overline{C}Y_t^{\geqslant} \setminus \underline{C}Y_t^{\geqslant} \tag{12}$$

$$BN_C(Y_t^{\leqslant}) = \overline{C}Y_t^{\leqslant} \setminus \underline{C}Y_t^{\leqslant} \tag{13}$$

We define the accuracy of approximation of $Y_t^{\geqslant}$ and $Y_t^{\leqslant}$, $t \in T$, respectively, as:

$$\alpha_C(Y_t^{\geqslant}) = \frac{|\underline{C}Y_t^{\geqslant}|}{|\overline{C}Y_t^{\geqslant}|} \tag{14}$$

$$\alpha_C(Y_t^{\leqslant}) = \frac{|\underline{C}Y_t^{\leqslant}|}{|\overline{C}Y_t^{\leqslant}|} \tag{15}$$

The ratio

$$\gamma_C(Y) = \frac{\left|U\left(\left(\bigcup_{t\in T} BN_C(Y_t^{\geqslant})\right) \cup \left(\bigcup_{t\in T} BN_C(Y_t^{\leqslant})\right)\right)\right|}{|U|} \tag{16}$$

defines the quality of approximation of the partition Y by means of the set of criteria C. This express the relation between all C-correctly classified objects and all the the object in the decision table.

Now we will introduce a formal language to describe approximation in logical terms.

With every set of criteria $Q \subseteq C \cup \{d\}$ we can associate a formal langage, i.e., a set of formulas $For(Q)$. Formulas of $For(Q)$ are built up from criterion-value pairs (q, v), where $q \in Q$, $v \in V$ by means of logical connectives $\wedge$ (and), $\vee \sim$ (or), (not) in the standard way.

For any $\varphi \in For(Q)$ by $\|\varphi\|_S$ we denote the set of all objects $x \in U$ satisfying φ in S and refer to as the meaning of φ in S.

The meaning $\|\varphi\|_S$ of φ in S is defined inductively as follows:

$$\|(q, v)\|_S = \{x \in U : q(x) = v\},\ q \in Q,\ v \in V_q \tag{17}$$

$$\|\varphi \vee \phi\|_S = \|\varphi\|_S \cup \|\phi\|_S \tag{18}$$

$$\|\varphi \wedge \phi\|_S = \|\varphi\|_S \cap \|\phi\|_S \tag{19}$$

$$\|\sim\varphi\|_S = U \backslash \|\varphi\|_S \tag{20}$$

A decision rule in S is an expression $\varphi \to \phi$, read "if φ then ϕ", where $\varphi \in For(C)$, $\phi \in For(d)$; φ and ϕ are referred to as conditions and decisions of the rule, respectively.

A decision rule $\varphi \to \phi$ is true in S, if $\|\varphi\|_S \subseteq \|\varphi\|_S$.

The number $supp_S(\varphi, \phi) = \text{card}(\|\varphi \wedge \phi\|_S)$ will be called the support of the rule $\varphi \to \phi$ in S. We consider a probability distribution $p_U(x) = \frac{1}{\text{card}(U)}$, $x \in U$, where U is non empty universe of objects in S; we have $p_U(X) = \frac{\text{card}(X)}{\text{card}(U)}$, $X \subseteq U$. For any formula φ we associate its probability in S, defined by $\pi_S(\varphi) = p_U(\|\varphi\|_S)$.

With every decision rule $\varphi \rightarrow \phi$ we associate a conditional probability $\pi_S(\phi \mid \varphi) = p_U(\|\phi\|_S \mid \|\varphi\|_S)$ that ϕ is true in S given φ is true in S called certainty factor, used first by J. Łukasiewicz to estimate the probability of implications. We have

$$\pi_S(\phi \mid \varphi) = \frac{\mathrm{card}(\|\varphi \wedge \phi\|_S)}{\mathrm{card}(\|\varphi\|_S)}, \quad \text{where} \quad \|\varphi\|_S \neq 0 \tag{21}$$

This coefficient is now widely used in data mining and called confidence coefficient. Obviously, $\pi_S(\phi \mid \varphi) = 1$, if and only if $\varphi \rightarrow \phi$ is true in S.

Beside, we will also use a coverage factor (used e.g. by T. Tsumato for estimation of the quality of decision rules) defined by

$$\pi_S(\varphi \mid \phi) = \frac{\mathrm{card}(\varphi \wedge \phi\|_S)}{\mathrm{card}(\|\phi\|_S)} \tag{22}$$

The certainty factor in S can be interpreted as the frequency of objects having the property ϕ in the set of objects having the property φ and the coverage factor – as the frequency of objects having the property φ in the set of objects having the property ϕ.

All the information about the algorithmic details of rules induction process are presented in J. W. Grzymała-Busse (1992), J. Stefanowski (2001); the concepts are used in "4eMka" system which is made by P. Słowiński and his team from Poznań University of Technology (http://www.idss.cs.put.poznan.pl/4eMka/index.html).

3. THE MULTICRITERIA EU'S COUNTRIES CLASSIFICATION PROBLEM

Presently, in the age of the European Union creation process, one of the most important problem is to answer the questions about the economic characteristics of EU'members, and then those classification into the groups. The aim of the analysis is to extract the knowledge about the main economic indicators and those values through the years which characterize each of the EU'countries.

The real data about the economics characteristics for the EU's countries are provided from the Internet stream (http://europa.eu.int/comm/economy--finance/publications/statistical-en.htm, ECFIN/174/2004-EN). Each of the countries is evaluated by six main economic indicators (as pointed out by J. Osiewalski and A. Welfe (1999)), i.e.: Export of goods and services (called Export), Import of goods and services (Import), Total factor productivity growth (called Productivity), Employment rate (called Employment), Unemployment rate (called Unemployment) and Nominal wages per head (called Wages).

The information about the main economic indicators is provided by the Eurostat Department (the definitions about the indicators are available in Internet stream (http://europa.eu.int/comm/economy-finance/publications/statistical-en.htm,ECFIN/174/2004-EN).

The evaluation of the EU'coutries was conduced through the years:

1) Italy, France, Luxemburg, Belgium, Germany, Netherlands, Denmark, Ireland, United Kingdom, Greece, Spain, Portugal, Austria, Finland, Sweden are evaluated in the following stages: 1961–1970, 1971–1985, 1986–1990, 1991–1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003,

2) Poland, Czech Republic, Estonia, Cyprus, Latvia, Lithuania, Hungary, Malta, Slovakia, Slovenia, Bulgaria, Romania, Turkey are evaluated in the following stages: 1992, 1993, 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003.

The decision table contains 336 objects (28 countries multiply by 12 stages of evaluation).

The *a priori* partition of the EU'countries is the foundation of the rough sets approach to decision model design. The a priori partition contains seven classes, presents the historical context of EU'extension process and simultaneously reflects the increase preferences of economic importance of related countries i.e.:

- class Y_7 (1952 – the beginning of EU): Italy, France, Luxemburg, Belgium, Germany,
- class Y_6 (1973 – the first extension of EU): Denmark, Ireland, United Kingdom,
- class Y_5 (1981 – the second extension of EU): Greece,
- class Y_4 (1986 – the third extension of EU): Spain, Portugal,
- class Y_3 (1995 – the fourth extension of EU): Austria, Finland, Sweden,
- class Y_2 (2004 – the fifth extension of EU): Poland, Czech Republic, Estonia, Cyprus, Latvia, Lithuania, Hungary, Malta, Slovakia, Slovenia,
- class Y_1 (a future extension of EU): Bulgaria, Romania, Turkey.

In order to explain the decision table one of its part is presented in Tab. 1.

Table 1

A part of decision table

Country	Year	Ex	Im	Prod	Emp	Unemp	Wg	*d*
France	1961–70	9.1	10.4	3.3	58.6	2.0	9.9	7
...	...	...	...	...	...	...	...	...
Poland	1999	–2.6	1	1.5	58.6	13.4	23.2	2
Poland	2000	23.2	15.6	58.6	58.6	16.4	1.8	2
...	...	...	...	...	...	...	...	...
Turkey	2003	16	27.1	5	58.6	10.8	31.4	1

Source: own calculations.

Then, it is clear, that the objects rule in the decision table plays the countries evaluated through the years by the main economic indicators, and the criteria rule plays the main economic indicators.

The aim of the analysis is to discover the knowledge about the economic indicators and its values through the years which in the best way characterize the EU's countries. Also, by the created decision model, we can control the economic growth level in the future years. It permit us to answer the question about the moment and the character of change in the economic situation, especially for the younger EU's members, e.g. using the criteria values for Poland for 2004 year, we can know the economic growth level is seriously increased, and since 2004 Poland could be assigned to higher class of countries, or not.

The results of rough sets analysis:

• accuracy of approximations for cumulated decision classes:

$\alpha_C(Y_2^{\geqslant}) = 0.51$, $\alpha_C(Y_3^{\geqslant}) = 0.39$, $\alpha_C(Y_4^{\geqslant}) = 0.27$, $\alpha_C(Y_5^{\geqslant}) = 0.27$, $\alpha_C(Y_6^{\geqslant}) = 0.25$, $\alpha_C(Y_7^{\geqslant}) = 0.12$,

$\alpha_C(Y_1^{\leqslant}) = 0.07$, $\alpha_C(Y_2^{\leqslant}) = 0.33$, $\alpha_C(Y_3^{\leqslant}) = 0.32$, $\alpha_C(Y_4^{\leqslant}) = 0.46$, $\alpha_C(Y_5^{\leqslant}) = 0.49$, $\alpha_C(Y_6^{\leqslant}) = 0.5$.

Quality of classification: $\gamma_C(Y) = 0.26$;

• the decision model contains 153 decision rules. Some of the strongest ones for related cumulative decision classes with corresponding probabilistic characteristics are presented bellow:

Rule 1: (Productivity ($\leqslant -7.4$) $\rightarrow d \leqslant 1$.

Support: Turkey 1994, Turkey 1999, Turkey 2001. The support indicate quite univocally that, by the total productivity growth, the situation in Turkey through the indicated years was the same.

$\pi_S(\phi \mid \varphi) = 0.75$ – 75% of countries having the properties "Productivity $\leqslant -7.4$" belongs to the cumulated class $Y_1^{\leqslant} = Y_1$,

$\pi_S(\varphi | \phi) = 0.23$ – 23% of countries from cumulated class $Y_1^{\leqslant}$ satisfies the condition "Productivity $\leqslant -7.4$".

Rule 9: (Import $\geqslant 27.1) \rightarrow d \leqslant 2$.

Support: Poland 1996, Czech Republic 1992, Latvia 1996, Estonia 1997, Estonia 2000, Cyprus 1995, Romania 1995, Romania 2000, Turkey 1993, Turkey 1995, Turkey 2003.

Looking on the above support and comparing it with the support of next rule we can conclude that the polish economic growth, almost by the import of goods and services, is increasing.

$\pi_S(\phi | \varphi) =$ – 100% of countries having the properties "Import $\geqslant 27.1$" belongs to the cumulated class $Y_2^{\leqslant}$,

$\pi_S(\varphi | \phi) = 0.14$ – 14% of countries from cumulated class $Y_2^{\leqslant}$ satisfies the condition "Import $\geqslant 27.1$".

Rule 40: (Import $\geqslant 16.9$) $\wedge$ (Unemployment $\geqslant 6.5) \rightarrow d \leqslant 3$.

Support: Spain 1986–1990, Finland 2000, Poland 1996, Poland 1997, Poland 1998, Czech Republic 1992, Czech Republic 1993, Czech Republic 2000, Slovakia 1996, Lithuania 1996, Lithuania 1997, Latvia 1998, Hungary 1993, Hungary 1998, Slovenia 1993, Bulgaria 2000, Romania 2001, Turkey 1993, Turkey 1995, Turkey 1997, Turkey 2000, Turkey 2003.

$\pi_S(\phi | \varphi) = 1$ – 100% of countries having the properties "(Import $\geqslant 16.9$) $\wedge$ (Unemployment $\geqslant 6.5$)" belongs to the cumulated class $Y_3^{\leqslant}$,

$\pi_S(\varphi | \phi) = 0.24$ – 24% of countries from cumulated class $Y_3^{\leqslant}$ satisfies the condition "(Import $\geqslant 16.9$) $\wedge$ (Unemployment $\geqslant 6.5$)".

Rule 48: (Import $\geqslant 11.5$) $\wedge$ (Export $\leqslant 13.8$) $\wedge$ (Wages $\leqslant 18.5$) $\wedge$ (Employment $\leqslant 65.8$) $\wedge$ (Productivity $\leqslant 5.7) \rightarrow d \leqslant 4$.

Support: France 1998, France 2000, Ireland 1996, Spain 1961–1973, Spain 1986–1990, Spain 1998, Spain 1999, Portugal 1961–1973, Portugal 1986–1990, Portugal 1998, Austria 1997, Austria 2000, Sweden 1997, Czech Republic 1992, Czech Republic 1994, Czech Republic 1996, Czech Republic 2001, Slovakia 1996, Slovakia 1998, Latvia 1998, Latvia 2001, Latvia 2003, Estonia 1998, Hungary 1999, Hungary 2003, Slovenia 1993, Slovenia 1994, Slovenia 1997, Bulgaria 2001, Bulgaria 2003, Romania 2001.

From the practical point of view there is not an interesting rule because of its small discriminatory power – the rule describe a mix of powers and not powers. There is important to remember that it could be a consequence of cumulating process of decision classes: $Y_1^{\leqslant}$, $Y_2^{\leqslant}$, $Y_3^{\leqslant}$ and $Y_4^{\leqslant}$. The higher value of $\pi_S(\phi | \varphi)$ is not surprised, then

$\pi_S(\phi | \varphi) = 0.97$,
$\pi_S(\varphi | \phi) = 0.22$.

Rule 66: (Unemployment $\geqslant 11.8$) $\wedge$ Export $\leqslant 19.2$) $\wedge$ (Import $\geqslant 4.9$) $\wedge$ (Productivity $\leqslant 2.2$) $\wedge$ (Employment $\leqslant 59.5) \rightarrow d \leqslant 5$.

Support: France 1997, Greece 1999, Spain 1986–1990, Spain 1991–1995, Spain 1996, Spain 1997, Spain 1998, Spain 1999, Poland 1993, Poland 1994, Poland 1996, Poland 2003, Slovakia 1995, Slovakia 1997, Slovakia 2000, Slovakia 2001, Slovakia 2002, Lithuania 1998, Lithuania 2003, Latvia 1997, Latvia 1998, Latvia 2000, Latvia 2001, Hungary 1993, Bulgaria 1997, Bulgaria 1999, Bulgaria 2000, Bulgaria 2001, Bulgaria 2003.

$\pi_S(\phi \mid \varphi) = 0.4,$
$\pi_S(\varphi \mid \phi) = 0.19.$

In order to test the quality of induced model author have reclassified all objects from learning set. The result: 46% correct answers.

In the last step the classification of the 28 analyzed countries for 2004 year have been conduced, e.g. Poland could be classified into the classes: $Y_2^{\leqslant}$, $Y_3^{\leqslant}$, $Y_4^{\leqslant}$, $Y_5^{\leqslant}$, $Y_6^{\leqslant}$, or $Y_7^{\leqslant}$. There isn't strong discrimination result, any way, our situation in the 2004 wasn't the worst, i.e. the rule brakes Poland classification into the worst class $Y_1^{\leqslant}$.

4. CONCLUSIONS

The created decision model based on dominance relation contains more general rules than in indiscernibility approach. For this reason and by the probabilistic characteristics easier permit to support the decision maker in the decision process about future EU's countries classification, then.

It is true that the quality of classification obtained by the indiscernibility or similarity approach is better then the above (on the similarity approach the highest quality of classification is obtained, e.g. I. Gruszka (2005)), but we need a dominance relation in order to solve a multiple criteria classification problems. A joint of indiscernibility and dominance approaches will be presented in future papers (like it is suggested e.g. in Greco, Materazzo, Słowiński 1996).

REFERENCES

Greco S., Materazzo B., Słowiński R. (1996a), *Rough Approximation of a Preference Relation by Dominance Relation*, ICS Research Report 16/96, **13**.

Greco S., Materazzo B., Słowiński R. (1996b), *On Joint Use of Indiscernibility, Similarity and Dominance Rough Approximation of Decision Classes*, ICS Research Report 16/96, **11**.

Greco S., Materazzo B., Słowiński R. (1999), *The Use of Rough Sets and Fuzzy Sets in MCDM*, [in:] *Multicriteria Decision Making: Advances in MCDM Models, Algorithms, Theory, and Applications*, Kluwer Academic Publihers, Boston–Dordrecht–London.

G r u s z k a I. (2005), *The Decision Analysis Based on Similarity Relation. An Application of The Rough Sets Theory to Multiattribute EU's Countries Classification Problem*, WISSC-4'04 Fourth Warsaw International Seminar on Soft Computing (in appear) **15**.

G r z y m a ł a - B u s s e J. W. (1992), *LERS – A System for Learning from Examples Based on Rough Sets*, [in:] *Intelligent Decision Support. Handbook of Applications and Advances of The Rough Sets Theory*, R. Słowinski (ed.), Kluwer Academic Publisher, **15**.

O s i e w a l s k i J., W e l f e A. (1999), *A Short-Run Price-Wage Nexus: An Application of Endogenous Switching*, "Przegląd Statystyczny", **11**, 4, 435–440.

P a w l a k Z. (1982), *Rough sets*, "International Journal of Computer and Information Sciences", **15**, 11.

P a w l a k Z. (1999), *Rough Classification*, "International Journal of Human Computer Studies", **51**, 14.

P a w l a k Z. (2001), *Rough Sets and Decision Algorithms*, [in:] *RSCTC 2000, LNAI 2005*, W. Ziarko, Y. Yao (eds.), Springer-Verlag, Berlin–Heidelberg.

S t e f a n o w s k i J. (2001), *Algorytmy indukcji reguł decyzyjnych w odkrywaniu wiedzy*, "Rozprawy", **361** (Wydawnictwo Politechniki Poznańskiej, Poznań).

Acknowledgements. This research has been supported by grant No. 2 HO2B 006 22 from State Committe for Research (Komitet Badań Naukowych).

Iwona Gruszka

ZASTOSOWANIE ZBIORÓW PRZYBLIŻONYCH DO WIELOKRYTERIALNEGO PROBLEMU KLASYFIKACJI PAŃSTW UNII EUROPEJSKIEJ W OPARCIU O RELACJĘ DOMINACJI. PROBABILISTYCZNE WŁAŚCIWOŚCI GENEROWANYCH REGUŁ DECYZYJNYCH

Teoria zbiorów przybliżonych została wprowadzona przez Z. P w l a k a w 1982. Matematyczną podstwą zstosowania zbiorów przybliżonych jest relacja binarna określona na uniwersum obiektów. W klasycznej analizie zbiorów przybliżonych jest to relacja nierozróżnialności. Jako relacja równoważności pozwala ona dzielić uniwersum obiektów na klasy równoważności, które stanowią atomy wiedzy o uniwersum. W celu wyindukowania dobrego modelu (z możliwie małą liczbą silnych reguł) niezbędny jest proces granulacji. Niemniej jednak z innego punktu widzenia całkiem naturalne wydaje się rozszerzenie koncepcji nierozróżnialności w celu rozważenia sytuacji, gdy jedne obiekty dominują nad innymi ze wyględu na rozważane kryteria, których zbiory wartości są uporządkowne zgodnie z preferencjami decydenta.

S. G r e c o, B. M a t e r a z z o i R. S ł o w i ń s k i (1999a, b, 1999) zaproponowali rozszerzenie teorii zbiorów przybliżonych – relacja nierozróżnialności została zastąpiona relacją dominacji.

Celem analizy przeprowadzonej przez autorkę jest indukcja modelu decyzyjnego i określenie probabilistycznych właściwości generowanych reguł decyzyjnych dla problemu wielokryterialnej klasyfikacji państw Unii Europejskiej. Analiza została przeprowadzona przez autorkę w systemie „4eMka".