Polish J. of Environ. Stud. Vol. 16, No. 4A (2007), 372-375

Short Communication

Application of Rough Sets for Identification of Factors Determining Sold Production of Industry

B. Wąsikowska

Institute of IT in Management, Faculty of Economics and Management, University of Szczecin, Mickiewicza 64, 71-101 Szczecin, Poland

Abstract

The Rough Set Theory has been framed by Professor Zdzisław Pawlak in 1982. It serves as a tool for analysing and reducing data sets. In particular, it has found application in data exploration, in solving complex classification problems and in computer support for various decision processes. Underneath, a method of identifying factors that significantly affect the **sold production of industry** as well as construction of its model using the rough sets are presented.

Keywords: computer, aided process planning systems

Introduction

At the turn of the 90s, a programme of sudden transition from the centrally planned economy to the market-controlled one was implemented in Poland. This was also a transition from a seller's market with long-lasting supply shortages to a buyer's market with demand barriers. During this transition programme, administrative price regulation was almost abolished. The money became completely freely exchangeable in compliance with the rate of exchange formed on the market. Therefore, a market succeeded to be developed, where a company and a household can buy all types of products and services without any limitations, both these manufactured within the country and those coming from the import. However, a by-product of all these moves was a considerable decline in industrial production and national income. Furthermore, production decline in companies resulted in a reduction of wages and an increase of unemployment, which in turn brought about a decrease in demand both for consumption produce of industrial origin and foodstuffs. This again resulted in a drop of income in the population living on agriculture (being almost a one third of the country's population), becoming thus an additional reason of fall in demand for products and services of non-agricultural origin. The abatement of this type of demand deepened the production decline in companies. This production decline in companies made worse their financial standing and reduced company demand for investments. In the mid-90s, the economic situation in Poland has started to improve and the level of the **sold production of industry** started to grow.

As it can be seen from the aforesaid, the production, being one of the essential elements of the Polish economy, is distinguished by a large vulnerability to changes in macroeconomic environment and many factors influence its volume. In this connection, it was determined to use the rough set theory for identifying the factors that significantly influence the level of the **sold production of industry.**

Identification of the Economic Factors by Means of Rough Sets

The first step in identifying the economic factors that influence the production and in constructing its model is

Table 1. Fragment of primary information table

U	Exports of commodities total	Investment outlays	Employed persons in enterprise sector	...	Population	Production
p1	0.00000	0.00000	0.98475	...	0.00000	0.08258
p2	0.07066	0.15316	0.77795	...	0.01937	0.00000
p3	0.23199	0.21319	0.83905	...	0.04263	0.27823
p4	0.12824	0.16369	0.81982	...	0.06976	0.15649
p5	0.11470	0.24551	0.65479	...	0.10852	0.14581
⋮	⋮	⋮	⋮	⋮	⋮	⋮
p93	0.86386	0.83421	0.36794	...	0.90310	0.86452
p94	1.00000	1.00000	0.20755	...	0.90310	0.91258
p95	0.92778	0.88694	0.33887	...	0.89147	0.80089
p96	0.78696	0.84293	0.27008	...	0.86434	0.75044

Source: own study.

presentation of collected statistical material in the form of so-called primary information table (Table 1) [1-3].

The next step consists in transformation of data contained in the information table into a coded form. In the case described below, a computer programme **Analysis of data using the rough sets**, working in the **Matlab 6.0** environment, was used for performing a discretisation of variables. The range of the explained variable was separated into three equal intervals, while ranges of the explaining variables were divided into two intervals, depending on the distribution of a given variable. As a result of the performed coding, a seconddary information table was received that contained twenty one conditional attributes (explaining variables) and one decision attribute (explained variable). The main objective of the analysis of data with the rough sets is to reduce maximally the variables and to determine a decision algorithm. This objective can be achieved by accomplishing the following stages [2, 3]:

1. determination of the elementary sets of examples,
2. determination of decision concepts,
3. reduction of the number of conditional attributes,
4. construction of a decision algorithm.

The Elementary Sets of Examples

When analysing the particular objects of secondary information table, one may observe that some of them contain identical values of conditional attributes. In the whole secondary table, sixty five so-called elementary sets were identified that contained indistinguishable objects with regard to the values of conditional attributes.

Decision Concepts

In the following step, the so-called decision concepts were determined, i.e. the sets that contained objects having the same decision value. Due to the fact that a decisive attribute may assume one of three different values (i.e. 1 or 2 or 3) the same will be also the number of decision concepts.

The decision concept X_1 contains all these objects (examples), for which d=1.

X_1={p1, p2, p3, p4, p5, p6, p7, p8, p10, p11, p13, p14, p16, p17, p18, p19, p26, p28, p38, p42, p43, p50, p73, p74, p85}

The decision concept X_2 contains all these objects (examples), for which d=2.

X_2={p9, p12, p15, p20, p21, p22, p23, p24, p25, p27, p29, p30, p31, p32, p36, p37, p39, p40, p41, p44, p45, p47, p48, p49, p51, p52, p53, p54, p55, p56, p59, p61, p62, p63, p64, p65, p66, p67, p68, p71, p72, p76, p77, p78, p79, p80, p86, p88, p91}

The decision concept X_3 contains all these objects (examples), for which d=3.

X_3={p33, p34, p35, p46, p57, p58, p60, p69, p70, p75, p81, p82, p83, p84, p87, p89, p90, p92, p93, p94, p95, p96}

Reduction of the Number of Conditional Attributes and Construction of a Decision Algorithm

In the next step, an attempt was undertaken to reduce the number of conditional attributes. As a result of the

Table 2. Fragment of a decision algorithm

Rules	Attributes												Decision	Rule support	Rule strength	Rule certainty	Examples supported a rule
	q1	q4	q6	q7	q9	q11	q13	q14	q16	q17	q18	q20	d				
R1	1	1	2	1	1	1	1	1	1	1	2	2	1	3	0,03	1	p1, p2, p13
R2	1	1	2	1	1	2	1	2	1	2	2	2	1	1	0,01	1	p3
R3	1	1	2	1	1	2	1	1	2	1	2	2	1	6	0,06	1	p4, p5, p8, p10, p11, p16
R4	1	1	2	1	1	1	1	2	1	2	2	1	1	1	0,01	1	p6
R5	1	1	2	1	1	2	1	1	2	1	2	1	1	1	0,01	1	p7
R6	1	1	2	1	1	2	1	1	2	2	2	2	2	4	0,04	0,67	p9, p15, p20, p21
R7	1	1	2	1	1	2	1	2	2	2	2	2	2	1	0,01	1	p12
R8	1	1	2	1	1	2	1	1	1	2	2	1	1	1	0,01	1	p14
R9	1	1	2	1	1	2	1	1	2	2	2	2	1	2	0,02	0,33	p17, p18

Source: own study.

carried out examination, a set was received that contained twelve most important conditional attributes. In the rough set theory, this set called a reduct: $\{q_1, q_4, q_6, q_7, q_9, q_{11}, q_{13}, q_{14}, q_{16}, q_{17}, q_{18}, q_{20}\}$.

The attributes, which were not situated within the reduct, were removed from the decision table. So, there were removed the following attributes: q_2, q_3, q_5, q_8, q_{10}, q_{12}, q_{15}, q_{19}, q_{21}. Then, basing on the determined reduct, a decision algorithm was constructed that contained 68 rules (Table 2), including 14 conflicting ones with regard to decision attribute. In Table 2 are also presented the following numerical characteristics [2]:

- rule support = number of examples supporting a given rule,
- rule strength = rule support / number of examples (strength value was rounded up or down with 0.01 accuracy).
- rule certainty (value 1 means that a rule is 100% certain),
- examples (objects) were given that supported a given rule.

In the subsequent steps, an attempt was undertaken to simplify similar rules and to remove conflicting rules from the decision algorithm. As a result of the carried out simplifications, seven decision rules were received at last. Exemplary rule and its economic interpretation are presented below.

Exemplary rule:

If (q1=2) and (q4=2) and (q7=1) and (q9=2) and (q11=1) and (q13=2) and (q14=2) and (q16=2) and (q17=2) and (q18=1) and (q20=1) **then** (d=3).

Economic interpretation of exemplary rule:

If exports of commodities total ∈(0,3982-1> and average nominal wage and salary ∈ (0,5226-1> and unemployment benefits ∈ <0-0,4159> and investment outlays ∈ (0,2850-1> and receipts from personal income tax ∈ <0-0,6264> and income from total activity ∈ (0,2625-1> and cost of obtaining income from total activity ∈ (0,2280-1> and revenue of the state budget ∈ (0,4167-1> and expenditure of the state budget ∈ (0,3125-1> and price indices of sold production of industry ∈ <0-0,4054> and monthly inflation ∈ <0-0,3846> **then** value of sold production of industry ∈ (0,66-1>.

Conclusions

As it can be seen form the examination carried out above, the rough sets allow not only for determining which of the variables significantly influence the examined phenomenon but also for constructing the rules that describe relationships between variables occurring in the examined information system. An additional quality of the rough sets is that determined rules have their economic interpretation and that they can be presented in the linguistic form.

References

1. MRÓZEK A., PŁONKA L.: Analiza danych metodą zbiorów przybliżonych. Zastosowania w ekonomii, medycynie i sterowaniu, Akademicka Oficyna Wydawnicza PLJ, Warszawa **1999**.
2. PIEGAT A.: Zastosowania zbiorów przybliżonych w ekonomii, wykłady - rękopis powielony, Uniwersytet Szczeciński, Szczecin **2005/2006**.
3. RUTKOWSKI L.: Metody i techniki sztucznej inteligencji, PWN, Warszawa **2005**.