

dr Barbara Wąsikowska
Instytut Informatyki w Zarządzaniu
Uniwersytet Szczeciński

mgr Fatimah Furaiji
Computer Science Dept.
College of Science
University of Basrah

Examining male and female purchase preferences by means of rough set method

Introductory remarks

Preferences examining is currently the main field of researchers interest in economics and management. It is a main research category in modeling consumer behaviour on the market, which takes into consideration rationality of decision taking. The term 'preferences' is commonly used in scientific literature, however its meaning is different depending on a particular discipline. Preferences are a basic term in the economics theory, especially in the consumer choice theory. Usually, it is assumed that preferences reflect and formalise personal tastes and do not depend on economical factors (such as: goods price, consumer budget) but only on subjective feelings (such as satisfaction, contentment, happiness) or usefulness. Individual preferences allow making choice in a situation when there are many alternatives available.

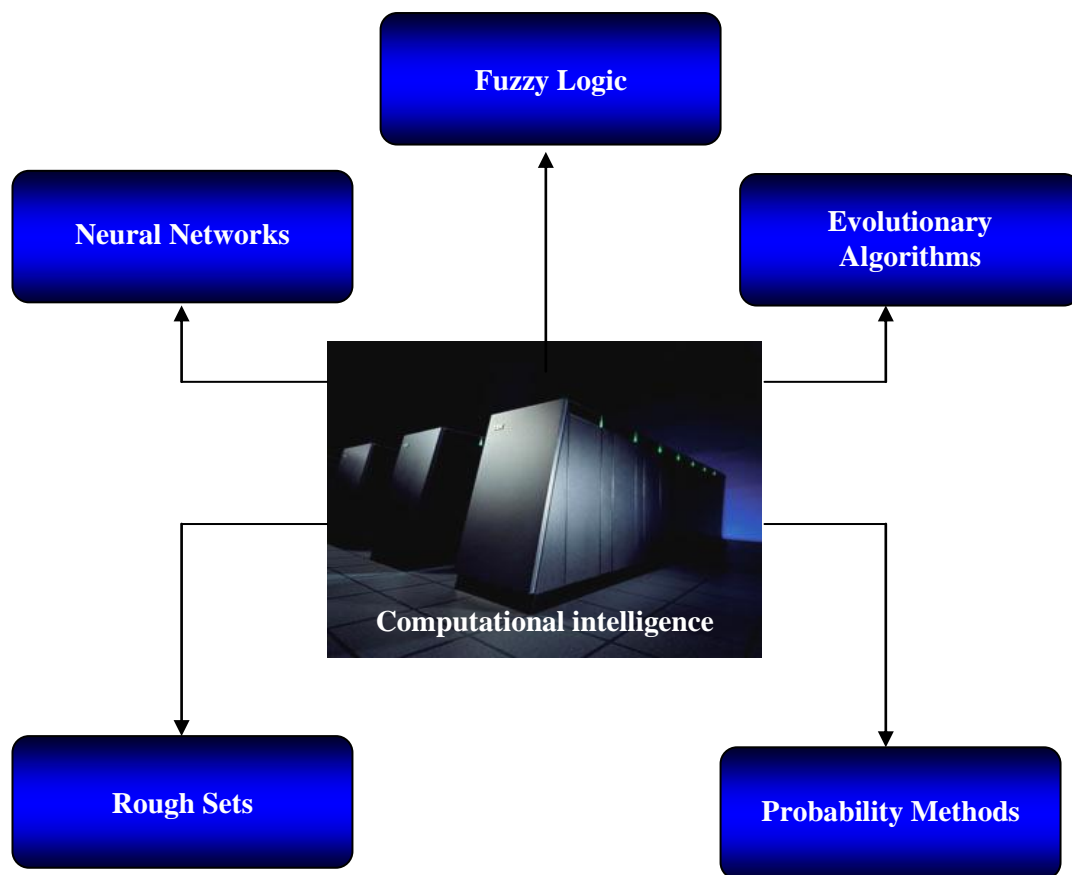
The aim of the article is the presentation of the way of marketing data analysis by an artificial intelligence method i.e. rough sets. Data, necessary to carry out the research, were collected using survey method in 2011 in Iraq and concern purchasing preferences of men and women. In the article authors verify the view that the application of the rough set theory to the analysis of marketing data may pose a significant supplement for cognitive examinations carried out using traditional methods.

Applying artificial intelligence methods in marketing research

During last few years there could be observed a dynamic development of artificial intelligence (IT discipline bordering with other disciplines). This particularly fast artificial intelligence development is tightly connected with the increase of computer power, as well as with increase of the amount of available data. The essence of systems based on artificial intelligence is the possibility of processing data expressed in numerical, binary, as well as linguistic or logical form. These data can also consist of orderly sequences of elements or tables and may contain elements described in a very inaccurate or even subjective way.

Until recent days the verbal information (eg. small company income, high product quality) was not used at all in methods based on conventional mathematics. Not taking these data into consideration considerably confined efficiency and effectiveness of different modeling, forecasting, market research, designing methods etc. Researchers started searching for new methods which would allow using these types of information as well. In recent years there have been created many hybrid methods joining learning systems with evolution and fuzzy systems. The basic methods of computational intelligence which emerged as a result of conducted research are shown in picture 1.

Picture 1. The basic methods of computational intelligence



Source: Author's own elaboration based on: Rutkowski L., Methods and techniques of artificial intelligence, PWN, Warsaw 2005.

Methods shown in picture 1. find many applications in business fields. One of such fields is marketing, particularly the area of communication with customer. It has a particular importance in a period of time when concepts of partnership marketing and creating long-term relationships with clients became extraordinarily popular (CRM). Contemporary software using methods and techniques of artificial intelligence can be

used in several areas connected with building long-term relationships with clients, for example:

- a) preparing clients database and cleaning it,
- b) conducting dialogue with the client in a natural language,
- c) automatic answering clients requests sent by e-mail.

Artificial intelligence methods are also more and more often used for examining consumer behaviour while shopping. Research development in the field of neurobiology and human brain functioning also allows knowing consumer behaviour better. The knowledge about buyer becomes more precise if it arises as a result of interdisciplinary view on the conditions of his actions¹.

Rough sets in defining rules of consumer behaviour

Rough set theory started by Zdzisław Pawlak in the 1980s and improved for a number of years by its inventor and many other researchers is a theoretical basis of many different IT methods and tools, which allow, among others, decision-making rules induction or data sets reduction. Rough set theory became highly popular and is currently used for data mining and knowledge discovering, complex classification tasks as well as computer Decision Support Systems. Disciplines in which this theory has been applied are, among others: medicine, pharmacology, business (banking, market research), speech and image recognition, linguistics, artificial intelligence. There is a number of factors which influence this method popularity. Firstly, rough set theory together with algorithms which use it, allow analysing large data sets enabling reduction of describing objects features sets to the essential minimum (so called, conditional attribute set reduct), removing data inconsistency, if such exists as well as generating from the data so called minimal rules, which means simulating particular discipline expert activity who would usually be able to present his knowledge as such rules (records in a form of : If... and ... so ...). The following should be perceived as an advantage of rough set theory:

- a) it does not require assumptions concerning the data (eg. probability or fuzziness),
- b) it contains fast data analysis algorithms,
- c) it facilitates the interpretation of the results,

¹ J. Żmija: *from the review of the book H. Mruk, M. Sznajder: Neuromarketing. Interdisciplinary view on the customer, Poznań Natural Science University Publishing House, Poznań 2008.*

d) it has considerable mathematic simplicity²

In the present paper there has been shown a way of rough sets application for creating a basis of men and women behaviour rules while buying electrical appliances. The question which the authors ask was: how important is for women and men high safety standards assurance of electrical appliances bought by them? The research data were collected by conducting a survey among man and women living in Basrah in Iraq. The data were then arranged in a form of so called primary information table (table 1). By information table we understand orderly five $T=(U, Q, D, V, f)$, where D is a set of decision features (attributes), U is a set of objects (examples) and is called universum, Q is a set of features (conditional attributes), $V=Uq \in Q$ i Vq is a set of all possible features value, while $f: U \times Q \rightarrow V$ is an informational function³. The table way of presenting data has got two basic features: universality, which means it allows gaining and keeping different kinds of data; and effectiveness, which allows easy computer analysis of data recorded in such way.

The first column contains numbers given to the particular surveyed persons (from p1 to p174). The next three columns contain conditional attribute values. These attributes are:

- sex of the surveyed person,
- age of the surveyed person,
- education of the surveyed person.

Table 1. Fragment of the primary information table

L.p.	Sex	Age	Education	I prefer electrical appliances with high safety standards
p1	Female	18-24	noneducation	disagree
p2	Male	more than 54	noneducation	fairly
p3	Male	more than 54	secondary	totally agree
p4	Male	35-44	high	totally agree
p5	Female	45-54	secondary	totally agree
p6	Female	18-24	high	totally agree
p7	Female	18-24	high	totally agree
p8	Female	18-24	secondary	totally agree
p9	Female	25-34	high	totally agree
p10	Female	25-34	high	totally agree

² Z. Pawlak: *Rough sets, new mathematical method of data analysis*, in: Warsaw Polytechnic Monthly Magazine No 5/2004, p. 2.

³ L. Rutkowski: *Methods and techniques of artificial intelligence*, PWN, Warsaw 2005, p. 21.

p11	Male	25-34	high	agree
p12	Male	more than 54	secondary	totally disagree
p13	Female	25-34	high	totally disagree
.
.
.
p174	Female	35-44	secondary	totally agree

Source: own elaboration.

In the last column there is a decision attribute informing about how important high safety standards assurance of the bought electrical appliances was to a particular surveyed person.

As a next step, in accordance with rough sets theory, the data aggregated in the primary information table were subjected to discretization, which means coding. The way of coding values of particular attributes is shown in Table 2.

Tabela 2. Discretization of attributes values of the primary information table.

The way of coding sex attribute	The way of coding age attribute	The way of coding education attribute	The way of coding decision attribute
Attribute symbol: q1 M – (male) F – (female)	Attribute symbol : q2 v1 – [from 18 to 24 years old] v2 – [from 25 to 34 years old] v3 – [from 35 to 44 years old] v4 – [from 45 to 54 years old] v5 – [from 55 and more]	Attribute symbol: q3 v1 – didn't get education v2 – primary v3 – medium v4 – secondary v5 – higher	Attribute symbol: d A1 – total disagree, disagree A2 – fairly A3 – agree, total agree

Source: own elaboration.

As a result of the conducted coding the secondary information table was received (table 3).

Table 3. Fragment of the secondary information table

No	q1 (sex)	q2 (age)	q3 (education)	d (decision)
p1	F	v1	v1	A1
p2	M	v5	v1	A2
p3	M	v5	v4	A3
p4	M	v3	v5	A3
p5	F	v4	v4	A3
p6	F	v1	v5	A3
p7	F	v1	v5	A3
p8	F	v1	v4	A3
p9	F	v2	v5	A3
p10	F	v2	v5	A3
.
.
.
p174	F	v3	v4	A3

Source: own elaboration

The data prepared in such way were analysed by rough set method using DAT (Data Analysis Toolbox) working in MATLAB R2010a environment. On the basis of initially conducted calculations, that is delimitation of elementary sets (sets of examples containing the same conditional attribute values) decision concepts (sets of examples containing the same decision value) there was an attempt to reduce conditional attribute set. As a result of the conducted research it was stated that all conditional attributes are important (none of the attributes is redundant) and they are the core of the set. The importance of particular conditional attributes was examined by calculating normalised importance ratio (table 4).

Table 4. Importance of particular conditional attributes

Attribute	Importance
Sex	0.2880
Age	0.6400
Education	0.4000

Source: own elaboration

The analysis of table 2 shows that importance of each of the chosen attributes is higher than zero, which confirms the fact that it is impossible to remove any of them from the decision table – all the attributes are more or less important.

The attribute is more important if the value of the attribute importance is closer to unity.

The most important two features are: age and education of the surveyed person.

Another step was the approximation of the set, which means calculating lower and upper approximations, set boundary as well as positive and negative regions (the exact explanation of these terms and appropriate formulas can be found in the following publications)⁴. On the basis of the received results approximation gamma quality ratio was calculated. It informs us of the percentage in which the analysed examples and conditional attributes allow defining certain rules. The quality of set F approximation is expressed by the following formula:

$$\gamma_{\tilde{Q}}(F) = \frac{\text{card}(\text{Pos}_{\tilde{Q}}(F))}{\text{card}(U)},$$

In the numerator there is the number of examples included in positive region and in the denominator there is the number of examples included in the whole set universum.

In this case, gamma ratio value is 72%, which means that 72% examples in the information table generates rules 100% certain. As a result of conducting the following analyses there was received decision table containing 42 rules. The fragment of the decision table is shown in table 5. There were also shown the following numeric characteristics:

- rule support which means the number of examples confirming a particular rule,
- rule certainty, which is calculated by dividing rule support/ number of rules having the same premise,
- there were specified examples (objects) supporting a particular rule.

This table was divided into well-defined part and badly-defined part. In the badly-defined part contradictory rules were included (total of 12 rules), while in the well-defined part other 30 rules were included.

The next step was to simplify similar rules which were in the well-defined part. Table 6 shows decision algorithm including exemplary received rules.

Table54. Fragment of the decision table

No.	Sex	Age	Education	Decision attribute	Rule support	Rule certainty	Rule supporting examples
1	F	v1	v1	A1	1	1	p1
2	F	v4	v3	A3	1	1	p95
3	F	v4	v4	A3	7	1	p5, p49, p75, p129, p163, p168, p170

⁴ Mrózek, L. Płonka: *Rough sets method data analysis. Applications in economics, medicine and steering*, Academic Publishing House PLJ, Warsaw 1999, p. 9 - 25.

Z. Pawlak: *Rough sets, new mathematical method of data analysis*, in: Warsaw Polytechnic Monthly Magazine No 5/2004, p. 4 - 5.

L. Rutkowski: *Methods and techniques of artificial intelligence*, PWN, Warsaw 2005, p. 20 - 45.

4	F	v4	v5	A1	1	0.1	p14
5	F	v4	v5	A3	9	0.9	p23, p44, p80, p118, p125, p138, p152, p153, p166
6	F	v5	v4	A3	1	1	p56
7	F	v5	v5	A3	3	1	p74, p164, p171
8	M	v1	v1	A3	1	1	p92
9	M	v1	v3	A3	1	1	p18
10	M	v1	v5	A3	4	1	p30, p87, p111, p148
11	M	v2	v2	A3	2	1	p80, p101
12	M	v2	v3	A3	5	1	p32, p38, p68, p72, p120
13	M	v2	v4	A3	4	1	p40, p93, p104, p150
14	M	v2	v5	A3	18	1	p11, p15, p20, p47, p48, p50, p53, p70, p99, p107, p108, p112, p116, p129, p145, p149, p158, p161

Source: own elaboration

Table 6. Decision algorithm

Rule number	Exemplary rules of decision algorithm
1	If sex = F and age = v1 and education = v1 is d=A1
2	If sex = F and age = v3 and education = v5 is d=A3
3	If sex = M and age = v2 and education = v4 is d=A3
4	If sex = M and age = v5 and education = v5 is q5 = A1

Source: own elaboration

After analysing rules from the well-defined part also contradictory rules were analysed. The contradictory rules should not be rejected beforehand. They require decent analysis. It can turn out that information contained in them might be very useful. Despite the fact that these rules are contradictory, attention should be paid to the number of cases confirming a particular rule. For instance, rule No 3 is supported by 14 cases in 16 including the same conditional attribute values (the characteristics called rule certainty informs about this). Meanwhile, rules contradictory to it (rule 1 and 2) are supported by only one case. It is worth considering if the rightness of the rule 3 should be taken into consideration and rules 1 and 2 - rejected as incorrect. Similar issue occurs in case of other contradictory rules included in the table 7.

Table 7. Chosen contradictory rules

Rule number	Examples of contradictory rules	Rule certainty
1	If sex =F and age = v2 and education = v5 to d=A1	1/ 16
2	If sex =F and age = v2 and education = v5 to d=A2	1/ 16
3	If sex =F and age = v2 and education = v5 to d=A3	14/ 16
4	If sex =F and age = v4 and education = v5 to d=A1	1/ 10
5	If sex =F and age = v4 and education = v5 to d=A3	9/ 10
6	If sex =M and age = v3 and education = v4 to d=A1	1/ 10
7	If sex =M and age = v3 and education = v4 to d=A3	9/ 10

Source: own elaboration

Finally, as a result of conducted simplification of the rules, eight rules were received . Among these 8 rules, two of them deserve particular attention. They are rules confirmed by a very large number of cases. (Table 8).

Table 8. Rules confirmed by the largest number of cases.

Rule number	Rule	Number of casus confirming the rule
1	If sex = woman and education \neq lack then decision = important or very important. Verbal description of the rule: For a women having at least primary education electrical appliances high safety standards assurance is important or very important.	66
2	If sex = man and age = above 24 and education \neq lack then decision = important or very important. Verbal description of the rule: For a man, aged 24 and above, having at least primary education, electrical appliances high safety standards assurance is important or very important.	90

Source: own elaboration

Conclusions on the conducted research

On the basis of the received rules and the analysis of the importance of particular conditional attributes it has been stated that attributes which has the greatest influence on the decision attribute are:

- a) for women: education,
- b) for men, except education also age is an important factor.

Of the received rules, supported by the largest number of cases, it clearly comes out that for women having minimum primary education, independently from age, electrical appliances high safety standards assurance is important or very important.

Whereas for men, such assurance is important only for men older than 24. For younger men caring about electrical appliances safety was not so important. For people not having any education electrical appliances high safety standards assurance was not important at all. An interesting fact is that among men having higher education were such for whom the problem of electrical appliances safety was not important, while among educated women were not any such cases.

Summary

The research presented in the article was focused on the data analysis concerning purchase preferences of men and women. The aim of the research was to check if while buying electrical appliances the surveyed respondents paid attention to high safety standards of the bough goods.

To analyse the collected data rough sets were used. The received results allow conclusion that the used method of artificial intelligence i.e. rough set method can be successfully used in practice as an effective tool for this type of data analysis .

The created basis of purchase preferences rules for men and women can be used as a base of knowledge for companies producing electrical appliances and can be a direction showing what this two groups of consumers pay attention to while buying appliances produced by these companies.

Examining male and female purchase preferences by means of rough set method

Summary

The paper presents nonstandard way of conducting research on male and female purchase preferences. The main emphasis is put on the method which was used in the research – the rough set theory. This method was applied to identify rules of male and female behavior while buying electrical appliances.

Keywords: consumer behaviuor, preferences, rough set theory, data mining.