DISCOVERING KNOWLEDGE WITH THE ROUGH SET APPROACH

Mazurek J.*

Abstract: The rough set theory, which originated in the early 1980s, provides an alternative approach to the fuzzy set theory, when dealing with uncertainty, vagueness or inconsistence often encountered in real-world situations. The fundamental premise of the rough set theory is that every object of the universe is associated with some information, which is frequently imprecise and insufficient to distinguish among objects. In the rough set theory, this information about objects is represented by an information system (decision table). From an information system many useful facts and decision rules can be extracted, which is referred as knowledge discovery, and it is successfully applied in many fields including data mining, artificial intelligence learning or financial investment. The aim of the article is to show how hidden knowledge in the real-world data can be discovered within the rough set theory framework. After a brief preview of the rough set theory's basic concepts, knowledge discovery is demonstrated on an example of baby car seats evaluation. For a decision rule extraction, the procedure of Ziarko and Shan is used.

Keywords:
information system, knowledge discovery,[3] rough sets, rule extraction, uncertainty.

Introduction

The rough set theory was proposed by a Polish computer scientist Zdislaw I. Pawlak in 1982; see e.g. [8], [9] or [10]. It is a mathematical tool for handling uncertainty and vagueness in decision making processes. The theory is based on an assumption that every object of the universe is associated with some information, such as price, quantity or durability in economics. However, some objects might be indiscernible when they are associated with the same information. That's why a set of such objects cannot be defined precisely (as a crisp set), and is formally approximated by rough sets – a pair of sets which give its lower and upper approximation.
Since 1980s, the rough set theory was successfully applied to many fields ranging from data mining to artificial intelligence learning. The main benefits of a rough sets model according to Tay and Shen [14]:

- It doesn't need any external information such as knowledge of probability distribution in statistics or a membership function in fuzzy set theory.
- It allows both for quantitative and qualitative analysis.
- It enables to discover fact hidden in a database and to express them as decision rules.

* **Jiří Mazurek**, Sileasian University in Opava, School of Business Administration in Karviná, Department of Mathematical Methods in Economics
✉ corresponding author: mazurek@opf.slu.cz.

- It eliminates redundant information of original data.
- The decision rules are supported by real examples contained in the data.
- Results of the rough set model are easy to understand and interpret.

In economics, rough sets models such as RSES, LERS, DataLogic, TRANCE or ProbRough are used for [14]:

- Business failure prediction, see e.g. [13] or [2],
- Database marketing, see e.g. [11],
- Financial investment, see e.g. [16].

After its introduction in early 1980s, the rough set theory was studied intensively by a large number of experts and was extended into (group) multicriteria decision analysis (see e.g. [4] or [5]), fuzzy sets ([3]), machine learning ([15]) and other fields of mathematics and computer science.

The aim of the article is to show how hidden knowledge in the real-world data can be discovered within the rough set theory framework. This might be helpful in manager's work, as it can facilitate understanding of data and information in general. The paper is organized as follows: Section 2 provides a brief preview of the rough set theory's basic concepts, in Section 3 an example – the evaluation of baby car seats – is analyzed within the rough set theory and in Section 4 rule extraction from an information system in Section 3 is demonstrated. Conclusions close the article.

A brief preview of the rough set theory

In the rough set theory, objects of analysis and their evaluation by multiple attributes is represented by an *information system*. An information system with two disjoint classes of attributes – *condition attributes C* and *decision attributes D* – is called a *decision table* (*U*, *C*, *D*), where:

- U is a nonempty, finite set of objects x.
- C is a nonempty, finite set of condition attributes, $C = \{C_1, C_2, ..., C_n\}$, and each attribute $C_i \in C$ is a function $C_i : U \to V(C_i)$, where $V(C_i)$ is a value set of C_i,
- D is a nonempty, finite set of decision attributes $D = \{d_1, d_2, ..., d_m\}$, and each attribute $d_i \in D$ is a function $d_i : U \to V(d_i)$, where $V(d_i)$ is a value set of d_i.

The *indiscernibility relation* R_B on U associated with a set $B \subseteq C \cup D$ is defined as:

$$xR_B y \Leftrightarrow a(x) = a(y); \forall a \in B; x, y \in U$$

A set of objects indiscernible with an object x by a relation R_B is called an *equivalent class* $[x]_B$. Every indiscernibility relation provides a *partition* of U into equivalent classes.

In the rough set theory, a set X cannot be expressed exactly (in general), as some elements from U might be indiscernible by a set of attributes B (see Example 1) Let $X \subseteq U$ and $B \subseteq C \cup D$, then $\underline{B}(X)$ and $\overline{B}(X)$ denote *lower* and *upper approximation* of a set X with respect to a relation R_B , and are defined as:

$$\underline{B}(X) = \{x \in U; [x]_B \subseteq X\}$$

$$\overline{B}(X) = \{x \in U; [x]_B \cap X = \varnothing\}$$

According to definitions above, the lower approximation (a *positive region*) of a set X is a union of all classes which are subsets of X (are contained in X), thus objects in $\underline{B}(X)$ *positively* (surely) belong to a set X.

The upper approximation of the set X is a union of all classes which have nonempty intersection with X, thus objects in $\overline{B}(X)$ can *possibly* be members of X.

A tuple $(\underline{B}X, \overline{B}X)$ is called a *rough set*, which means that a rough set is represented by two crisp sets – its *lower* and *upper boundary*. A set $BN_B(X) = \overline{B}X - \underline{B}X$ is called a *boundary region* of a set X, and it contains objects that cannot be ruled in or out as members of a set X. When $BN_B(X) = \varnothing$, X is a crisp set, otherwise it is a rough set with respect to B.

The accuracy of rough set approximation of a set X induced by an indiscernibility relation R_B is given as:

$$\alpha_B(X) = \frac{|\underline{B}X|}{|\overline{B}X|}$$

where $|\ |$ denotes a cardinality of a given set. Clearly, $\alpha_B(X) \in [0,1]$. When the lower approximation is equal to the upper approximation, then $\alpha_B(X) = 1$ and the approximation is perfect. When the lower approximation is an empty set, the accuracy is zero. A number $\rho_B(X) = 1 - \alpha_B(X)$ is a *roughness* of a set X.

A *reduct* (RED) is a subset of a set of attributes, which fully characterizes the information system; hence it is a sufficient set of attributes to represent information system’s structure. A reduct satisfies two conditions:

i) $[x]_{RED} = [x]_{C \cup D}$, so the equivalence classes generated by a reduct are the same as equivalence classes generated by a set of all attributes.

ii) The RED is a minimal set of attributes which satisfies i).

A reduct is not a unique set in general. A set of attributes belonging to all reducts is called a *core*. Attributes in core are indispensable; they cannot be removed from an

information system without loss of information in a system, while other attributes are dispensable.
In database analysis, another important question is attribute dependency. In the rough set theory, dependence γ of an attribute sets B on an attribute set A is given as:

$$\gamma(A,B) = \frac{\sum_{i=1}^{n} |\underline{A}B_i|}{|U|}$$

For each equivalent class $[x]_B$, its lower approximations by attributes in A are summed up and divided by a cardinality of a set U. Dependence is bounded: $0 \leq \gamma(A,B) = \leq 1$, and larger values denote stronger dependence B on A. It should be noted that $\gamma(A,B)$ is not symmetric.

The example: baby car seat evaluation

To illustrate some of the concepts above, consider an information system shown in Table 1, which describes the evaluation of baby car seats from a newspaper *Mladá Fronta Dnes* dated 13. 6. 2006 [7]. In the evaluation, nine car seats (labeled as A, B, C, ..., I) were assigned from 0 to 3 stars (the more stars the better evaluation) according to four criteria: *safety*, *handling*, *comfort* and *maintenance* (see Table 1). The overall classification was: *good*, *satisfactory* and *unsatisfactory*. Therefore, we have:

- the set of objects (car seats) U = {A, B, ... , I},
- the set of four condition attributes: $C = \{C_1, C_2, C_3, C_4\}$, where C_1 = 'safety', C_2 = 'handling', C_3 ='comfort' and C_4 ='maintenance',
- The decision attribute d = 'classification'.

Table 1. The evaluation of baby car seats

Car seat	safety	handling	comfort	maintenance	classification
A	3	3	2	3	good
B	3	3	3	2	good
C	3	3	3	2	good
D	3	3	3	3	good
E	3	3	3	2	good
F	3	3	3	2	good
G	2	3	2	3	satisfactory
H	0	3	3	2	unsatisfactory
I	0	2	2	2	unsatisfactory

Source: [7].

Let $B = \{C_1, C_2, C_3, C_4\}$ be a set of all condition attributes. With the use of the indiscernibility relation R_B we get the following family of equivalent classes:

{B, C, E, F}, {A}, {D}, {G}, {H}, {I}

Thus, objects B, C, E and F are indiscernible with regard to the relation R_B.
Now, consider a set $X = \{A, B, C, D\}$. Clearly, X cannot be expressed precisely, because we cannot distinguish among objects B, C, E and F. The lower approximation of the set X: $\underline{B}(X) = \{A, D\}$ and the upper approximation of the set X: $\overline{B}(X) = \{A, B, C, D, E, F\}$.
The accuracy of the rough set approximation of the set X by the relation R_B:

$$\alpha_B(X) = \frac{|\underline{B}X|}{|\overline{B}X|} = \frac{1}{3}$$

Information system in Table 1 consists of four condition attributes, bur not all of them are indispensable, because the reduct $RED = \{C_1, C_3, C_4\}$; hence, the condition attribute C_2 ('handling') is dispensable. As the reduct is unique, the core is equal to the reduct.
Even more intriguing feature of the rough set theory is its ability of finding hidden knowledge in the data. This problem is addressed in the next section.

Rule extraction

From a decision table a set of decision rules in the following form can be induced for each $x \in U$ (for each row of a table):

IF ($C_1, C_2, \ldots, C_n$) THEN ($d_1, d_2, \ldots, d_m$).

Each decision rule can be described by following properties [8]:

- *Support*: $supp_x(C,D)$, which is a number of objects x from U which satisfy the rule.
- *Strength*: $\sigma_x(C,D) = \frac{supp_x(C,D)}{|U|}$,
- *Certainty factor*: $cer_x(C,D) = \frac{supp_x(C,D)}{|C|}$. If $cer_x(C,D) = 1$, then a rule is called a *certain decision rule*, otherwise it is an *uncertain decision rule*.
- *Coverage factor*: $cov_x(C,D) = \frac{supp_x(C,D)}{|D|}$

For example, Table 1 contains following decision rule (see rows 2, 3, 5 and 6):

IF ($C_1 = 3$, $C_2 = 3$, $C_3 = 3$, $C_4 = 2$) THEN ($d = good$)

The decision rule has these properties: *supp* $(C, D) = 4$, $\sigma(C, D) = 4/9$, *cer* $(C, D) = 1$ and *cov* $(C, D) = 2/3$.

As each row of a decision table is associated with some decision rule, finding all *minimal rules* from a given information system, that is rules with a minimum number of rule conditions, which express the strongest patterns in the data, is of more interest (see e.g. [1] [6] or [12]). One of the most used models for minimal rules extraction is LERS (*Learning from Examples based on Rough Sets*) by Grzymala-Busse [6], which can induce rules even from inconsistent data, that is from the data were some object are evaluated equally by all condition attributes, but differently by at least one of decision attributes.

In this section a procedure proposed by Ziarko and Shan [17] will be used to illustrate knowledge discovery in the data with the information system in Table 1. In Ziarko and Shan approach, a decision matrix for each value of a decision attribute is formed in the first step, see Table 2 (d = 'good'), Table 3 (d = 'satisfactory') and Table 4 (d = 'unsatisfactory').

In a decision matrix represented by Table 2, objects with d = 'good' are placed in the rows while objects with other values are placed into columns. At rows' and columns' intersections differences among given objects are listed in the form of values C_j^i. For example, value C_1^3 at the intersection of objects B and H indicates, that the object B differs from the object H with regard to the condition attribute C_1 (and no other), and a value of this attribute for the object B is 3. The same procedure is applied in Table 3 and Table 4. As objects B, C, E and F are of the same class (they are indiscernible), only object B is shown.

Table 2. A decision matrix for the decision attribute value 'good'

Object	G	H	I
A	C_1^3	C_1^3, C_3^2, C_4^3	C_1^3, C_2^3, C_4^3
B	C_1^3, C_3^3, C_4^2	C_1^3	C_1^3, C_2^3, C_3^3
C	C_1^3, C_3^3, C_4^2	C_1^3	C_1^3, C_2^3, C_3^3
D	C_1^3, C_3^3	C_1^3, C_4^3	$C_1^3, C_2^3, C_3^3\ C_4^3$
E	C_1^3, C_3^3, C_4^2	C_1^3	C_1^3, C_2^3, C_3^3
F	C_1^3, C_3^3, C_4^2	C_1^3	C_1^3, C_2^3, C_3^3

Table 3. A decision matrix for the decision attribute value 'satisfactory'

Object	A	B	D	H	I
G	C_1^2	C_1^2, C_3^2, C_4^3	C_1^2, C_3^2	C_1^2, C_3^2, C_4^3	C_1^2, C_2^3, C_4^3

Table 4. A decision matrix for the decision attribute value 'unsatisfactory'

Object	A	B	D	G
H	C_1^0, C_3^3, C_4^2	C_1^0	C_1^0, C_4^2	C_1^0, C_3^3, C_4^2
I	C_1^0, C_2^2, C_4^2	C_1^0, C_2^2, C_3^2	$C_1^0, C_2^2, C_3^2, C_4^2$	C_1^0, C_2^2, C_4^2

In the second step, from a decision table Boolean expressions are formed for every row of a table. From Table 2 we get (while omitting rows 3, 5 and 6 identical with the 2nd row):

1st row:

$$(C_1 = 3) \wedge (C_1 = 3 \vee C_3 = 2 \vee C_4 = 3) \wedge (C_1 = 3 \vee C_2 = 3 \vee C_4 = 3) \Rightarrow d = good$$

2nd row:

$$[(C_1 = 3) \vee (C_3 = 2) \vee (C_4 = 2)] \wedge (C_1 = 3) \wedge [(C_1 = 3) \vee (C_2 = 3) \vee (C_3 = 3)] \Rightarrow d = good$$

4th row:

$$[(C_1 = 3) \vee (C_3 = 3)] \wedge [(C_1 = 3) \vee (C_4 = 3)] \wedge [(C_1 = 3) \vee (C_2 = 3) \vee (C_3 = 3) \vee (C_4 = 3)] \Rightarrow d = good$$

Simplifying expressions above, we obtain the following decision rules:

i) $(C_1 = 3) \Rightarrow d = good$,

ii) $(C_3 = 3) \wedge (C_4 = 3) \Rightarrow d = good$.

After repeating the procedure with the data in decision Tables 3 and 4, we obtain additional decision rules:

iii) $(C_1 = 2) \Rightarrow d = satisfactory$,

iv) $(C_1 = 0) \vee (C_2 = 2) \Rightarrow d = unsatisfactory$,

v) $(C_3 = 2) \wedge (C_4 = 2) \Rightarrow d = unsatisfactory$.

These results can be summarized as follows:

- To be classified *good*, a car seat have to be evaluated 3 stars in a category 'safety', or it must be evaluated 3 stars both in categories 'comfort' and 'maintenance'.
- A car seat is classified *satisfactory*, if it is given 2 stars in the category 'safety'.
- A car seat is classified *unsatisfactory*, if it is assigned 0 stars in categories 'safety' or 'handling'; or it is given 2 stars in both in categories 'comfort' and 'maintenance'.

Moreover, from the previous section we know that:

- Car seats formed six equivalent classes with regard to the set of all condition attributes: {B, C, E, F}, {A}, {D}, {G}, {H}, {I}.
- The category 'handling' was found redundant in the evaluation, as it was not included in the reduct.

Decision rules derived from a decision matrix allows the classification of a new object. In our example with baby car seats, when a hypothetical baby car seat is evaluated 2 stars in the category 'safety', then it follows from the 2nd decision rule that it would be classified as 'satisfactory'.

Summary

The aim of the article was to show how hidden knowledge in the real-world data can be discovered within the rough set theory framework. For knowledge discovery the approach of Ziarko and Shan was applied to the baby car seat evaluation with four condition attributes (safety, handling, comfort and maintenance) and three decision attributes (good, satisfactory and unsatisfactory) presented in the newspaper *Mladá Fronta Dnes*. From decision matrices decision rules listed in the previous section were extracted, furthermore, it was learned that one of condition attributes, namely handling, was redundant. This example demonstrated that the rough set approach can be useful also in a management, as knowledge acquisition is an important part of manager's work.

References:

[1]. Bazan, J. G. A comparison of dynamic and non-dynamic rough set methods for extracting laws from decision table. In: *Rough Sets in Knowledge Discovery*, 1998, vol. 1, Physica-Verlag, heidelberg, pp. 321-365.

[2]. Bazan, J. G., Skowron, A., Synak, A. Market data analysis: A rough set approach. *ICS Research Reports*, 1994, 6/94, Warsaw.

[3]. Dubois, D., Prade, H. Rough fuzzy sets and fuzzy rough sets. *International Journal of General Systems*, 1990, nr. 17, pp. 191-209.

[4]. Greco, S., Matarazzo, B., Slowinski, R. Rough sets theory for multicriteria decision analysis. *European Journal of Operational Research*, 2001, Vol. 129, Nr. 1, pp. 1–47.

[5]. Greco, S., Matarazzo, B., Slowinski, R. Rough set approach to multi-attribute choice and ranking problems. In: *Multiple Critetia Decision Making: 12th International Conference Proc*., Springer, Berlin: 1997.

[6]. Grzymala-Busse, J. A new version of the rule induction system LERS. *Fundamenta Informaticae*, 1997, Nr. 31.

[7]. Mazurek J., "*The Evaluation of Conflict's Degree in Group Decision Making",* Polish Journal of Management Studies, vol. 5/2012.

[8]. *Mladá Fronta Dnes*. Available from WWW: <http://auto.idnes.cz/bezpeci-pro-deti-test-autosedacek-dlq-/automoto.asp?c=A060612_202359_automoto_fdv>

[9]. Pawlak, I. Z. Rough set theory and its applications. *Journal of Telecommunications and Information Technology*, 2002, Nr. 3, pp. 7-10.

[10]. Pawlak, I. Z. Rough sets. *International Journal of Computer and Information Sciences*, 1982, Vol. 11, Nr. 5, pp. 341-356.

[11]. Pawlak, I. Z. *Rough sets: Theoretical Aspects of Reasoning About Data*. Kluwer Academic Publishers, Dordrecht: 1991.
[12]. Poel, D. Rough sets for database marketing. In: *Rough Sets in Knowledge Discovery*, Physica - Verlag, Wurzburg, pp. 324-335.
[13]. Skowron, A. Boolean Reasoning for decision rules generation. In: *Methodologies for Intelligent Systems, Lecture Notes in Artificial Intelligence*, Vol. 689, Springer, Berlin, pp. 295-305.
[14]. Slowinski, R., Zopounidis, C., Dimitras, A. I., Susmaga, R. Rough set predictor of business failure. In: *Soft Computing in Financial Engineering*, 1999, Physica-Verlag, Wurzburg, pp. 402-424.
[15]. Tay, F. E. H., Shen, L. Economic and financial prediction using rough sets model. *European Journal of Operations Research*, 2002, Vol. 141, pp. 641-659.
[16]. Wenshan, W., Haihua, L. Machine Learning Applications in Rough Set Theory. In: *Internet Technology and Applications*, 2010, Wuhan, pp. 1-3.
[17]. Ziarko, W., Golan, R., Edwards, D. An application of datalogic/R knowledge discovery tool to identify strong predictive rules in stock market data. In: *Proceedings of AAAI Workshop on Knowledge Discovery in Databases*, Washington, DC, 1993, pp. 89-101.
[18]. Ziarko, W., Shan, N. Discovering attribute relationships, dependencies and rules by using rough sets. In: *Proceedings of the 28th Annual Hawaii International Conference on System Sciences*, 1995, Hawaii, pp. 293-299.

ODKRYWANIE WIEDZY W PODEJŚCIU TEORII ZBIORÓW PRZYBLIŻONYCH

Streszczenie: Teoria zbiorów przybliżonych, która powstała w roku 1980, oferuje alternatywne podejście do teorii zbiorów rozmytych, gdy ma się do czynienia ze zjawiskiem niepewności, niejasności i niekonsekwencji, często spotykanym w rzeczywistych sytuacjach. Podstawowym założeniem teorii zbiorów przybliżonych jest to, że każdy obiekt wszechświata jest związany z pewnymi informacjami, które są często nieprecyzyjne i niewystarczające do rozróżnienia między obiektami. W teorii zbiorów przybliżonych, informacje o obiektach są reprezentowane przez system informacyjny (tabela decyzyjna). System informacyjny dostarcza wiele przydatnych faktów i reguł, które są określane jako odkrywanie wiedzy, która z powodzeniem jest stosowana w wielu dziedzinach, w tym w ekstrakcji danych, sztucznej inteligencji czy przy inwestycjach finansowych. Cele artykułu jest pokazanie, w jaki sposób wiedza ukryta w rzeczywistych danych, mogą zostać odkryte w trudnych ramach teorii mnogości. Po krótkim przedstawieniu podstawowych pojęć teorii zbiorów przybliżonych, na przykładzie ocen fotelików samochodowych, przedstawiono zjawisko odkrywania wiedzy. W celu wydobycia reguły decyzyjnej zastosowano procedurę Ziarko i Shan.

Słowa kluczowe: System informacyjny, odkrywanie Widzy, zbiory przybliżone, ekstrakcja zasad, niepewność

發現知識與粗糙集方法

摘要：粗糙集理論，它起源於20世紀80年代初，提供了一種替代的方法模糊集理論，當處理中經常遇到的現實世界的情況下，不確定性，模糊或不一致。粗糙集理論的基本前提是宇宙的每一個對象相關的一些信息，這是經常不精確，不足以區分對象。在粗糙集理論，這個對象有關的信息所代表的信息系統（決策表）。從信息系統中提取許多有用的事實和決策規則，這被稱為知識發現，它已成功應用於諸多領域，包括數據挖掘，人工智能學習或金融投資。本文的目的是展示如何在現實世界中的數據可以發現隱性知識粗糙集理論的框架內。經過簡短的預覽粗糙集理論的基本概念，知識發現是證明嬰兒汽車座椅評價的一個例子。對於決策規則提取，程序的Ziarko和擇使用