Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 39 (2019) 260–269

Transportation Research Procedia

www.elsevier.com/locate/procedia

Green Cities 2018

Usage of the rough set theory for generating decision rules of number of traffic vehicles

Marek Landowski[a]* Anna Landowska[b]

[a]Department of Mathematical Methods, Maritime University of Szczecin, Wały Chrobrego 1-2, 70-500 Szczecin, Poland
[b]Faculty of Management and Economics of Services, University of Szczecin, Cukrowa 8 Str., 70-001 Szczecin, Poland

Abstract

Often, it is difficult to interpret and use the large size of data obtained from the experiment. In addition, the generated information can be unprecise. The rough set theory besides probability theory, fuzzy set theory and many others in recent years is very often used by scientists to solve problems of data mining. In the paper the data mining of the traffic vehicles with rough set theory was made. With this theory it was shown that it is possible to generate the decision rules of the number of vehicles at the specific points in the city. On the basis of 120 objects 16 well-defined linguistic decision rules were obtained.

Keywords: rou rough sets; decision rules; transport; vehicle traffic volume

1. Introduction

The rough set theory was formulated by Zdzislaw I. Pawlak in 1982 (Pawlak, 1982, 1991). The idea of the rough set theory is based on the fact that knowledge can be classified. Every living being operating in the environment behaves in such a way that real or abstract objects (e.g. things or signals received by the senses) that surround it are classified in a different ways (Rutkowski, 2005). The classification capability is based on visible differences between objects. On the basis of this operation, classes of indistinguishable objects are built, that is, objects that do not differ from each other in a noticeable manner are assigned to the appropriate class of objects. From the received classes of

* Corresponding author. Marek Landowski
E-mail address: m.landowski@am.szczecin.pl

10.1016/j.trpro.2019.06.028

indistinguishable objects you can build knowledge using rules. In this way, a part of the reality that surrounds us or a part of the abstract world is created.

Application of the rough set theory is very wide. For example in data mining (Chen, et.al., 2015, Zarandi & Kazemi, 2008, Grzymala-Busse, 2005, Kumar & Yadav, 2015), medicine (Paszek, Wakulicz-Deja, 2007, Durairaj, Sathyavathi, 2013, Øhrn, 1999, Ilczuk, et al., 2005), processing of massive data (Slezak, 2007, Yang, et al., 2010, Yun, 2014), and many more fields.

The paper presents the application of rough set theory to extract knowledge from experimental data obtained from the investigation of traffic intensity in selected locations in Szczecin. Data for investigation were taken from the experiment conducted in Szczecin (Poland) as a part of the Green and Sustainable Freight Transport Systems in Cities (GRASS) project (Kijewska, Iwan, 2016, Kijewska, Johansen, Iwan, 2016, Kijewska, Konicki, Iwan, 2016, Kijewska, et al., 2017). Each example (object) consists of the 4 variables: a category of the vehicles, a place where the experiment was carried out, speed of the vehicles and a number of vehicles on the specific day of the week. In the considered example on the basis of 120 objects the rules that show the number of vehicles are generated. Considered 120 objects were obtained from over 245,000 samples of the experiment. Therefore, the main task, under certain assumptions, using the decision rules obtained by rough set theory is to find the answer to the question "How many cars can be expected, under a given conditions?".

To determine the decision rules with rough set theory first some important theoretical foundations are introduced.

The data of the examples that are analyzed are presented in the form of a table, where each column is called an attribute and every row is called an object. Two classes of attributes are considered: condition attributes $C=\{c0,\dots,cn\}$ and decision attributes $D=\{d0,\dots dm\}$. An universe U is a set that is nonempty and finite. Every object x from the universe U will generate the decision rule $C_x \rightarrow D_x$.

An *elementary condition set Ei* ($i=\{0,\dots,k\}$, $k\leq n\}$ is a set of objects of universe U with equal condition attributes. *Elementary decision set Gj* ($j=\{0,\dots,l\}$, $l\leq m\}$ is a set of objects of universe U where decision attributes are equal.

A lower approximation LA(Gj) of the elementary decision set Gj for given attribution sets C, is a set of elementary condition sets Ei where each object from Ei belongs to the elementary decision set Gj:

$$LA(Gj) = \{Ei: Ei \subseteq Gj\} \tag{1}$$

An upper approximation UA(Gj) of the elementary decision set Gj for given attribution sets C, is a set of elementary condition sets Ei where at least one object of Ei belongs to the elementary decision set Gj.

$$UA(Gj) = \{Ei: Ei \cap Gj \neq \emptyset\} \tag{2}$$

Positive region is a sum of all lower approximations of the elementary decision sets for given attribution sets C:

$$pos(G_C) = \bigcup_{j=0}^{l} LA(Gj) \tag{3}$$

Subset C* of the set C is called *reduct* if $pos(G_C) = pos(G_{C^*})$.

Accuracy of approximation is a quotient of cardinality of elements of lower approximation and cardinality of elements of upper approximation (Del Giudice, et al., 2017):

$$accuracy = \frac{\sum_{j=0}^{l} card(LA(Gj))}{\sum_{j=0}^{l} card(UA(Gj))} = \frac{card\ (pos(G))}{\sum_{j=0}^{l} card(UA(Gj))} \tag{4}$$

where *card* is a cardinality of the set.

Quality of approximation is a quotient of cardinality of elements of lower approximation cardinality of rules (number of the objects in the universe U) (Del Giudice, et al., 2017).

$$quality = \frac{\sum_{j=0}^{l} card(LA(Gj))}{card(U)} \tag{5}$$

The *support* of the decision rule is a number of objects that confirm this decision rule (Pawlak, 2002):

$$supp_x(C,D) = card(C(x) \cap D(x)) \quad (6)$$

The *strength* of the decision rule is a quotient of the support of the decision rule and cardinality of rules (number of the objects in the universe *U*) (Pawlak, 2002):

$$stre_x(C,D) = \frac{supp_x(C,D)}{card(U)} \quad (7)$$

The *certainty factor* of the decision rule is a quotient of the support of the decision rule and cardinality of the decision rules with equal condition attributes and different decision attributes (Pawlak, 2002):

$$cer_x(C,D) = \frac{supp_x(C,D)}{card(C(x))} \quad (8)$$

2. Coding the attributes and determination of the elementary sets

In the considered example three condition attributes are used. This condition attributes have a significant impact on the number of vehicles. These attributes are as follows:

- A category of the vehicle,
- The place where the investigation was carried out,
- Vehicle speed.

Decision attributes is the number of vehicles in a given category that have passed through a given measurement point with a certain average speed during the day.

Each attribute is coded, which means that each attribute is assigned to the appropriate numerical value. Table 1 – 4 present condition and decision attributes in the uncoded and coded forms. Let us assume the following denotation:

- c1 – vehicle category,
- c2 – point where the investigation was carried out,
- c3 – speed of the vehicle,
- d – number of the vehicles per day.

Table 1. Condition attribute c1, vehicle category.

Input data	Coded form
Cat. 2	1
Cat. 3	2
Cat. 4	3

Table 2. Condition attribute c2, point where the investigation was carried out.

Input data	Coded form
Wojska Polskiego St. I – WPI	1
Wojska Polskiego St. II – WPII	2
Wojska Polskiego St. III – WPIII	3
Jagiellońska St. I – JagI	4
Jagiellońska St. II – JagII	5
Jagiellońska St. III – JagIII	6
Monte Cassino St. – MC	7
Rayskiego St. – Ray	8

Table 3. Condition attribute c3, speed of the vehicle.

Input data	Coded form
speed under 20 km/h	1
speed from 20 to 35 km/h	2
speed over 35 km/h	3

Table 4. Decision attribute *d*, number of the vehicles per day.

Input data	Coded form
A very small number of vehicles from 0 to 24	1
A small number of vehicles from 25 to 200	2
Average number of vehicles from 201 to 6000	3
A large number of vehicles from 6001 to 8000	4
A very large number of vehicles over 8001	5

The attributes have the following values: c1 = {1,2,3}, c2 = {1,2,3,4,5,6,7,8}, c3 = {1,2,3}, d = {1,2,3,4,5}. The basic input set of the objects obtained from the investigation contains 120 examples (rules).

2.1. Determination of the elementary condition sets

Determining elementary condition sets is a grouping examples (objects) on the basis of the values of attributes c1, c2 and c3 into sets. That is determining sets of elements with equal conditional attributes. Elementary condition sets *Ei* for 3 conditional attributes C = {c1,c2,c3} are as follows:

E0 = {1,2,3,4,5}; E1 = {6,7,8,9,10}; E2 = {11};
E3 = {12,13,14,15}; E4 = {16,17,18,19,20}; E5 = {21,22,23,24,25};
E6 = {26,27,28,29,30}; E7 = {31,32,33,34,35}; E8 = {36,37,38,39,40};
E9 = {41,42,43,44,45}; E10 = {46,50}; E11 = {47,48,49};
E12 = {51}; E13 = {52,53,54,55}; E14 = {56,57,58,59,60};
E15 = {61,62,64,65}; E16 = {63}; E17 = {66,68};
E18 = {67,69,70}; E19 = {71,72,73,74,75}; E20 = {76,77,78,79};
E21 = {80}; E22 = {81,82,84}; E23 = {83,85};
E24 = {86,90}; E25 = {87,88,89}; E26 = {91,93,94,95};
E27 = {92}; E28 = {96,97,99,100}; E29 = {98};
E30 = {101,102,103,104}; E31 = {105}; E32 = {106,107,108,109};
E33 = {110}; E34 = {111,112,113,114,115}; E35 = {116,119};
E36 = {117,118,120}.

Cardinalities of the elementary condition sets Ei are presented below:

card(E0)=5; card(E1)=5; card(E2)=1; card(E3)=4; card(E4)=5; card(E5)=5;
card(E6)=5; card(E7)=5; card(E8)=5; card(E9)=5; card(E10)=2; card(E11)=3;
card(E12)=1; card(E13)=4; card(E14)=5; card(E15)=4; card(E16)=1; card(E17)=2;
card(E18)=3; card(E19)=5; card(E20)=4; card(E21)=1; card(E22)=3; card(E23)=2;
card(E24)=2; card(E25)=3; card(E26)=4; card(E27)=1; card(E28)=4; card(E29)=1;
card(E30)=4; card(E31)=1; card(E32)=4; card(E33)=1; card(E34)=5; card(E35)=2;
card(E36)=3.

2.2. Determination of the elementary decision set

Each elementary decision set *Gj* is an elementary set of objects with equal values of the decision attribute *d*.

G0 = {72,73,74,106,112,113,...,120} – very small number of vehicles from 0 to 24 per day

G1 = {61,66,71,75,...,83,86,...,94,96,97,98,101,103,104,105,107,...,111} – a small number of vehicles from 25 to 200 per day

G2 = {17,19,...,27,30,...,60,62,...,65,67,...,70,84,85,95,99,100,102} – average number of vehicles from 201 to 6000 per day

G3 = {6,...,11,16,18,28,29} – a large number of vehicles from 6001 to 8000 per day
G4 = {1,...,5,12,...,15} – a very large number of vehicles over 8001 per day
Cardinalities of the elementary decision sets *Gj* are as follows:
card(G0) = 13
card(G1) = 33
card(G2) = 55
card(G3) = 10
card(G4) = 9

2.3. Lower approximation LA(Gj) of the elementary decision set Gj

Lower approximation (1) of the elementary decision sets *Gj* and their cardinality for conditional attributes C = {c1,c2,c3}:
For G0 = {72,73,74,106,112,113,...,120}
LA(G0) = E35 ∪ E36; card(LA(G0)) = 5
For G1 = {61,66,71,75,...,83,86,...,94,96,97,98,101,103,104,105,107,...,111}
LA(G1) = E20 ∪ E21 ∪ E24 ∪ E25 ∪ E27 ∪ E29 ∪ E31 ∪ E33; card(LA(G1)) = 14
For G2 = {17,19,...,27,30,...,60,62,...,65,67,...,70,84,85,95,99,100,102}
LA(G2) = E5 ∪ E7 ∪ E8 ∪ E9 ∪ E10 ∪ E11 ∪ E12 ∪ E13 ∪ E14 ∪ E16 ∪ E18; card(LA(G2)) = 44
For G3 = {6,...,11,16,18,28,29}
LA(G3) = E1 ∪ E2; card(LA(G3)) = 6
For G4 = {1,...,5,12,...,15}
LA(G4) = E0 ∪ E3; card(LA(G4)) = 9
Positive region for 3 condition atributes {c1,c2,c3} with formula (3) equals:
pos(G_C) = E0 ∪ E3 ∪ E1 ∪ E2 ∪ E5 ∪ E7 ∪ E8 ∪ E9 ∪ E10 ∪ E11 ∪ E12 ∪ E13 ∪ E14 ∪ E16 ∪ E18 ∪ E20 ∪ E21 ∪ E24 ∪ E25 ∪ E27 ∪ E29 ∪ E31 ∪ E33 ∪ E35 ∪ E36
card(pos(G_C)) = 78
Quality of approximation, formula (5), of the condition attributes Gj of C = {c1,c2,c3} equals:

$$quality = card(pos(G_C)) / card(U) = 78/120 = 0.65.$$

So 65% of objects allow to generate certain rules.

2.4. Upper approximation LA(Gj) of the elementary decision set Gj

Upper approximation of the elementary decision sets Gj are calculated with formula (2). Upper approximations of Gj and their cardinality for the set of condition attributes C = {c1,c2,c3} are as follows:
For G0 = {72,73,74,106,112,113,...,120}
UA(G0) = E19 ∪ E32 ∪ E34 ∪ E35 ∪ E36; card(UA(G0)) = 19
For G1 = {61,66,71,75,...,83,86,...,94,96,97,98,101,103,104,105,107,...,111}
UA(G1) = E15 ∪ E17 ∪ E19 ∪ E20 ∪ E21 ∪ E22 ∪ E23 ∪ E24 ∪ E25 ∪ E26 ∪ E27 ∪ E28 ∪ E29 ∪ E30 ∪ E31 ∪ E32 ∪ E33 ∪ E34; card(UA(G1)) = 51
For G2 = {17,19,...,27,30,...,60,62,...,65,67,...,70,84,85,95,99,100,102}
UA(G2) = E4 ∪ E5 ∪ E6 ∪ E7 ∪ E8 ∪ E9 ∪ E10 ∪ E11 ∪ E12 ∪ E13 ∪ E14 ∪ E15 ∪ E16 ∪ E17 ∪ E18 ∪ E22 ∪ E23 ∪ E26 ∪ E28 ∪ E30; card(UA(G2)) = 62
For G3 = {6,...,11,16,18,28,29}
UA(G3) = E1 ∪ E2 ∪ E4 ∪ E6; card(UA(G3)) = 16
For G4 = {1,...,5,12,...,15}
UA(G4) = E0 ∪ E3; card(UA(G4)) = 9
Accuracy of approximation, formula (4), of the family of elementary decision sets Gj equals:

$$accuracy = card(pos(G_C)) / \sum_{j=0}^{l} card(UA(Gj)) = 78/(19+51+62+16+9) = 78/157 \approx 0.5.$$

3. Attempt to reduce the condition attributes

In this section it will be checked if the condition attributes set C1 = {c2,c3} is a reduct of the set the condition attributes set C = {c1,c2,c3}.

For the condition attributes set C1 = {c2,c3} the elementary condition sets are in the form as it was shown below:

E0 = {1,2,3,4,5,41,42,43,44,45,83,85}
E1 = {6,7,8,9,10,47,48,49,87,88,89}
E2 = {11,51,91,93,94,95}
E3 = {12,13,14,15,52,53,54,55,92}
E4 = {16,17,18,19,20,56,57,58,59,60,96,97,99,100}
E5 = {21,22,23,24,25,61,62,64,65,105}
E6 = {26,27,28,29,30,67,69,70,110}
E7 = {31,32,33,34,35,71,72,73,74,75,111,112,113,114,115}
E8 = {36,37,38,39,40,80,117,118,120}
E9 = {46,50,86,90}
E10 = {63,101,102,103,104}
E11 = {66,68,106,107,108,109}
E12 = {76,77,78,79,116,119}
E13 = {81,82,84}
E14 = {98}

Determination of the lower approximation of the elementary decision sets Gj for C1 = {c2,c3}:

For G0 = {72,73,74,106,112,113,...,120}
LA(G0) = ∅; card(*LA*(G0)) = 0
For G1 = {61,66,71,75,...,83,86,...,94,96,97,98,101,103,104,105,107,...,111}
LA(G1) =;∅; card(*LA*(G1)) = 0
For G2 = {17,19,...,27,30,...,60,62,...,65,67,...,70,84,85,95,99,100,102}
LA(G2) = ∅; card(*LA*(G2)) = 0
For G3 = {6,...,11,16,18,28,29}
LA(G3) = ∅; card(*LA*(G3)) = 0
For G4 = {1,...,5,12,...,15}
LA(G4) = ∅; card(*LA*(G4)) = 0

Conclusions:

- Set of condition attributes C1 = {c2,c3} is not a reduct of the set C = {c1,c2,c3}, because family of the elementary condition sets are not equal. Family C* consists of 37 elementary condition sets, but C1* has 15 elementary condition sets.
- Positive region of G_{C1} (formula (5)) of condition attributes {c2,c3} is empty set pos(G_{C1}) = ∅, card(pos(G_{C1}))=0. Quality of approximation of the family C1* of elementary decision sets Gj for C1 = {c2,c3} equals *quality* = 0/120 = 0. The attribute set C1 = {c2,c3} does not generate certain rules, while for C = {c1,c2,c3} quality of approximation equals 0.65.
- Condition attributes c1 cannot be removed from the condition attribute set C = {c1,c2,c3}.

Similarly, it can be shown that condition attributes c2 and c3 cannot be deleted from the condition attribute set C= {c1,c2,c3}.

4. Development of decision rules

Full set of decision rules can be obtained from combination identical values of decision attributes d and condition attributes c1, c2 and c3. Support, certainty and strength by formulas (6) – (8) of not simplified rules are presented in Table 5 (accuracy 0.01), calculations were made on the basis of 120 examples when typing manuscripts:

Table 5. Support, certainty and strength of not simplified rules.

	Rule	c1 vehicle cat.	c2 location	c3 speed	d number of vehicles	support	certaint y	strengt h	Equality for condition attr.
1	R0	1	1	3	5	5	1	0.04	
2	R1	1	2	3	4	5	1	0.04	
3	R2	1	3	2	4	1	1	0.01	
4	R3	1	3	3	5	4	1	0.03	
5	R4	1	4	2	4	2	0.4	0.02	R4, R5
6	R5	1	4	2	3	3	0.6	0.03	R4, R5
7	R6	1	5	2	3	5	1	0.04	
8	R7	1	6	2	3	3	0.6	0.03	R7, R8
9	R8	1	6	2	4	2	0.4	0.02	R7, R8
10	R9	1	7	1	3	5	1	0.04	
11	R10	1	8	2	3	5	1	0.04	
12	R11	2	1	3	3	5	1	0.04	
13	R12	2	2	2	3	2	1	0.02	
14	R13	2	2	3	3	3	1	0.03	
15	R14	2	3	2	3	1	1	0.01	
16	R15	2	3	3	3	4	1	0.03	
17	R16	2	4	2	3	5	1	0.04	
18	R17	2	5	2	2	1	0.25	0.01	R17, R18
19	R18	2	5	2	3	3	0.75	0.03	R17, R18
20	R19	2	5	1	3	1	1	0.01	
21	R20	2	6	1	2	1	0.5	0.01	R20, R22
22	R21	2	6	2	3	3	1	0.03	
23	R22	2	6	1	3	1	0.5	0.01	R20, R22
24	R23	2	7	1	2	2	0.4	0.02	R23, R24
25	R24	2	7	1	1	3	0.6	0.03	R23, R24
26	R25	2	8	1	2	4	1	0.03	
27	R26	2	8	2	2	1	1	0.01	
28	R27	3	1	2	2	2	0.67	0.02	R27, R29
29	R28	3	1	3	2	1	0.5	0.01	R28, R30
30	R29	3	1	2	3	1	0.33	0.01	R27, R29
31	R30	3	1	3	3	1	0.5	0.01	R28, R30
32	R31	3	2	2	2	2	1	0.02	
33	R32	3	2	3	2	3	1	0.03	
34	R33	3	3	2	2	3	0.75	0.03	R33, R35
35	R34	3	3	3	2	1	1	0.01	
36	R35	3	3	2	3	1	0.25	0.01	R33, R35
37	R36	3	4	2	2	2	0.5	0.02	R36, R38
38	R37	3	4	1	2	1	1	0.01	
39	R38	3	4	2	3	2	0.5	0.02	R36, R38
40	R39	3	5	1	2	3	0.75	0.03	R39, R40
41	R40	3	5	1	3	1	0.25	0.01	R39, R40
42	R41	3	5	2	2	1	1	0.01	
43	R42	3	6	1	1	1	0.25	0.01	R42, R43
44	R43	3	6	1	2	3	0.75	0.03	R42, R43
45	R44	3	6	2	2	1	1	0.01	
46	R45	3	7	1	2	1	0.2	0.01	R45, R46
47	R46	3	7	1	1	4	0.8	0.03	R45, R46
48	R47	3	8	1	1	2	1	0.02	
49	R48	3	8	2	1	3	1	0.03	

4.1. The division of the decision rule set into a well-defined and ill-defined part

A part of well-defined decision rules contained in the lower approximations of elementary decision sets Gj is presented in Table 6.

Table 6. Well-defined decision rules.

	Rule	c1 vehicle cat.	c2 location	c3 speed	d number of vehicles	support
1	R0	1	1	3	5	5
2	R1	1	2	3	4	5

3	R2	1	3	2	4	1
4	R3	1	3	3	5	4
5	R6	1	5	2	3	5
6	R9	1	7	1	3	5
7	R10	1	8	2	3	5
8	R11	2	1	3	3	5
9	R12	2	2	2	3	2
10	R13	2	2	3	3	3
11	R14	2	3	2	3	1
12	R15	2	3	3	3	4
13	R16	2	4	2	3	5
14	R19	2	5	1	3	1
15	R21	2	6	2	3	3
16	R25	2	8	1	2	4
17	R26	2	8	2	2	1
18	R31	3	2	2	2	2
19	R32	3	2	3	2	3
20	R34	3	3	3	2	1
21	R37	3	4	1	2	1
22	R41	3	5	2	2	1
23	R44	3	6	2	2	1
24	R47	3	8	1	1	2
25	R48	3	8	2	1	3

Ill-defined decision rules and the value of their certainty are presented in Table 7, contradictory rules of the full set of decision rules. Let us consider rule as useful if its certainty is greater than 0.5

Table 7. Ill-defined decision rules and the value of their certainty.

	Rule	c1 vehicle cat.	c2 location	c3 speed	d number of vehicles	support	certainty	the usefulness of the rule
1	R4	1	4	2	4	2	0.4	not useful
2	R5	1	4	2	3	3	0.6	useful
3	R7	1	6	2	3	3	0.6	useful
4	R8	1	6	2	4	2	0.4	not useful
5	R17	2	5	2	2	1	0.25	not useful
6	R18	2	5	2	3	3	0.75	useful
7	R20	2	6	1	2	1	0.5	not useful
8	R22	2	6	1	3	1	0.5	not useful
9	R23	2	7	1	2	2	0.4	not useful
10	R24	2	7	1	1	3	0.6	useful
11	R27	3	1	2	2	2	0.67	useful
12	R28	3	1	3	2	1	0.5	not useful
13	R29	3	1	2	3	1	0.33	not useful
14	R30	3	1	3	3	1	0.5	not useful
15	R33	3	3	2	2	3	0.75	useful
16	R35	3	3	2	3	1	0.25	not useful
17	R36	3	4	2	2	2	0.5	not useful
18	R38	3	4	2	3	2	0.5	not useful
19	R39	3	5	1	2	3	0.75	useful
20	R40	3	5	1	3	1	0.25	not useful
21	R42	3	6	1	1	1	0.25	not useful
22	R43	3	6	1	2	3	0.75	useful
23	R45	3	7	1	2	1	0.2	not useful
24	R46	3	7	1	1	4	0.8	useful

Table 8 presents decision rules well defined after the connection operation and useful ill-defined rules as well as their support, certainty and strength.

Table 8. Support, certainty and strength of well-defined and useful ill-defined rules after connection operation.

	Rule	c1 vehicle cat.	c2 location	c3 speed	d number of vehicles	suppor t	certaint y	strength
1	R0/3	1	1 OR 3	3	5	9	1	0.08

2	R1	1	2	3	4	5	1	0.04
3	R2	1	3	2	4	1	1	0.01
4	R6/10	1	5 OR 8	2	3	10	1	0.08
5	R9	1	7	1	3	5	1	0.04
6	R11	2	1	3	3	5	1	0.04
7	R12/13	2	2	NOT 1	3	5	1	0.04
8	R14/15	2	3	NOT 1	3	5	1	0.04
9	R16/21	2	4 OR 6	2	3	5	1	0.04
10	R19	2	5	1	3	1	1	0.01
11	R25/26	2	8	NOT 3	2	7	1	0.06
12	R31/32	3	2	NOT 1	2	3	1	0.03
13	R34	3	3	3	2	1	1	0.01
14	R37	3	4	1	2	1	1	0.01
15	R41/44	3	5 OR 6	2	2	2	1	0.02
16	R47/48	3	8	NOT 3	1	5	1	0.04
17	R5	1	4	2	3	3	0.6	0.03
18	R7	1	6	2	3	3	0.6	0.03
19	R18	2	5	2	3	3	0.75	0.03
20	R24	2	7	1	1	3	0.6	0.03
21	R27	3	1	2	2	2	0.67	0.02
22	R33	3	3	2	2	3	0.75	0.03
23	R39	3	5	1	2	3	0.75	0.03
24	R43	3	6	1	2	3	0.75	0.03
25	R46	3	7	1	1	4	0.8	0.03

4.2. Linguistic form of useful rules

Knowledge extraction coefficient (generalizations of examples) = number of examples / number of rules:

Knowledge extraction coefficient = 120 / 25 = 4.8

The knowledge extraction ratio equals 4.8, so on average 5 examples is described by one useful rule.

Table 9 presents linguistic form of well-defined decision rules (16 rules) and 9 ill-defined but useful rules as well as their certainty.

Table 9. Linguistic form of defined decision rules.

Rules	Linguistic form of the decision rule	certainty
R0/3:	IF (veh.cat.2) AND (location WPI OR WPIII) AND (speed over 35 km/h) THEN (a very large number of vehicles)	1
R1:	IF (veh.cat.2) AND (location WPII) AND (speed over 35 km/h) THEN (a large number of vehicles)	1
R2:	IF (veh.cat.2) AND (location WPIII) AND (speed from 20 to 35 km/h) THEN (a large number of vehicles)	1
R6/10:	IF (veh.cat.2) AND (location JagII OR Ray) AND (speed from 20 to 35 km/h) THEN (average number of vehicles)	1
R9:	IF (veh.cat.2) AND (location MC) AND (speed under 20 km/h) THEN (average number of vehicles)	1
R11:	IF (veh.cat.3) AND (location WPI) AND (speed over 35 km/h) THEN (average number of vehicles)	1
R12/13:	IF (veh.cat.3) AND (location WPII) AND (speed over 20 km/h) THEN (average number of vehicles)	1
R14/15:	IF (veh.cat.3) AND (location WPIII) AND (speed over 20 km/h) THEN (average number of vehicles)	1
R16/21:	IF (veh.cat.3) AND (location JagI OR JagIII) AND (speed from 20 to 35 km/h) THEN (average number of vehicles)	1
R19:	IF (veh.cat.3) AND (location JagII) AND (speed under 20 km/h) THEN (average number of vehicles)	1
R25/26:	IF (veh.cat.3) AND (location Ray) AND (speed under 35 km/h) THEN (a small number of vehicles)	1
R31/32:	IF (veh.cat.4) AND (location WPII) AND (speed over 20 km/h) THEN (a small number of vehicles)	1
R34:	IF (veh.cat.4) AND (location WPIII) AND (speed over 35 km/h) THEN (a small number of vehicles)	1
R37:	IF (veh.cat.4) AND (location JagI) AND (speed under 20 km/h) THEN (a small number of vehicles)	1
R41/44:	IF (veh.cat.4) AND (location JagII OR JagIII) AND (speed from 20 to 35 km/h) THEN (a small number of vehicles)	1
R47/48:	IF (veh.cat.4) AND (location Ray) AND (speed under 35 km/h) THEN (a very small number of vehicles)	1
R5:	IF (veh.cat.2) AND (location JagI) AND (speed from 20 to 35 km/h) THEN (average number of vehicles)	0.6
R7:	IF (veh.cat.2) AND (location JagIII) AND (speed from 20 to 35 km/h) THEN (average number of vehicles)	0.6
R18:	IF (veh.cat.3) AND (location JagII) AND (speed from 20 to 35 km/h) THEN (average number of vehicles)	0.75
R24:	IF (veh.cat.3) AND (location MC) AND (speed under 20 km/h) THEN (a very small number of vehicles)	0.6
R27:	IF (veh.cat.4) AND (location WPI) AND (speed from 20 to 35 km/h) THEN (a small number of vehicles)	0.67
R33:	IF (veh.cat.4) AND (location WPIII) AND (speed from 20 to 35 km/h) THEN (a small number of vehicles)	0.75
R39:	IF (veh.cat.4) AND (location JagII) AND (speed under 20 km/h) THEN (a small number of vehicles)	0.75
R43:	IF (veh.cat.4) AND (location JagIII) AND (speed under 20 km/h) THEN (a small number of vehicles)	0.75
R46:	IF (veh.cat.4) AND (location MC) AND (speed under 20 km/h) THEN (a very small number of vehicles)	0.8

5. Conclusions

The paper presents the use of rough set theory to analyze the problem of vehicle traffic volume. The generated rules show that it is possible to use the rough set theory to extract knowledge from data of the intensity of vehicle traffic. Based on 120 objects, 16 well-defined rules and 9 useful ill-defined rules were derived. In the analyzed example three condition attributes (category of the vehicles, place where the experiment was taken, speed of the vehicles) and one decision attribute (number of vehicles) are considered. The calculated indicator of the quality of approximation of the condition attributes shows that 65% of analyzed examples allow to generate certain rules. The knowledge extraction ratio of the rules obtained by rough set theory is 4.8, so on average 5 objects are described by one useful rule. It was also shown how to check the significance of conditional attributes. The proposed methodology can help reduce congestion in the city by predicting the intensity of the traffic

References

Chen, H., Li, T., Luo, C., Horng, S.-J., Wang, G., 2015. A Decision-Theoretic Rough Set Approach for Dynamic Data Mining. IEEE Transactions on Fuzzy Systems, 23, 1958–1970.

Del Giudice, V., De Paola, P., Cantisani, G.B., 2017. Rough Set Theory for Real Estate Appraisals: An Application to Directional District of Naples. Buildings, 7, 1–12.

Durairaj, M., Sathyavathi, T., 2013. Applying Rough Set Theory for Medical Informatics Data Analysis. International Journal of Scientific Research in Computer Science and Engineering, 1(5), 1–8.

Grzymala-Busse, J., 2005. Rough Set Theory with Applications to Data Mining. In: Gh. Negoita M., Reusch B. (eds) Real World Applications of Computational Intelligence. Studies in Fuzziness and Soft Computing, vol 179, Springer, Berlin, Heidelberg, 221–244.

Ilczuk, G., Mlynarski, R., Wakulicz-Deja, A., Drzewiecka, A., Kargul, W., 2005. Rough set techniques for medical diagnosis systems. Computers in Cardiology, DOI: 10.1109/CIC.2005.1588235.

Kijewska, K., Iwan, S., 2016. Analysis of the functioning of urban deliveries in the city centre and its environmental impact based on Szczecin example. Transportation Research Procedia, 12, 739–749.

Kijewska, K., Johansen, B. G., Iwan, S., 2016. Analysis of freight transport demand at Szczecin and Oslo area. Transportation Research Procedia, 14, 2900–2909.

Kijewska, K., Konicki, W., Iwan, S., 2016. Freight transport pollution propagation at urban areas based on Szczecin example. Transportation Research Procedia, 14, 1543–1552.

Kijewska, K., Iwan, S., Konicki, W., Kijewski, D., 2017. Assessment of freight transport flows in the city centre based on the Szczecin example - Methodological approach and results. Research in Transportation Business & Management, 24, 59–72.

Kumar, M., Yadav, N., 2015. Fuzzy Rough Sets and Its Application in Data Mining Field. Advances in Computer Science and Information Technology, 2(3), 237–240.

Øhrn, A., 1999. Discernibility and Rough Sets in Medicine: Tools and Applications. Norwegian University of Science and Technology Trondheim, Doctor of Engineering Thesis 999:133.

Paszek, P., Wakulicz–Deja, A., 2007. Applying Rough Set Theory to Medical Diagnosing. In: Kryszkiewicz M. et al. (eds) Rough Sets and Intelligent Systems Paradigms. RSEISP 2007. Lecture Notes in Computer Science, vol 4585, Springer, Berlin, Heidelberg, 427–435.

Pawlak, Z., 1982. Rough sets. International Journal of Parallel Programming 11 (5), 341–356.

Pawlak, Z., 1991. Rough Sets: Theoretical Aspects of Reasoning About Data. Kluwer Academic Publishing.

Pawlak, Z., 2002 Rough set theory and its application. Journal of Telecommunications and Information Technology, 3/2002, 7–10.

Rutkowski, L. 2005. Methods and techniques of artificial intelligence. Wydawnictwo Naukowe PWN, Warsaw.

Slezak, D., 2007. Rough Sets and Few-objects-many-attributes Problem: the Case Study of Analysis of Gene Expression Data Sets. Frontiers in the Convergence of Bioscience and Information Technologies, 437–442.

Yang, Y., Chen, Z., Liang, Z., Wang, G., 2010. Attribute Reduction for Massive Data Based on Rough Set Theory and MapReduce. In: Yu et al. (eds) Rough Set and Knowledge Technology 2010. Lecture Notes in Computer Science, vol 6401. Springer, Berlin, Heidelberg, 672–678.

Yun, S., 2014. Research of Big Data Analysis on Rough Set and Electromagnetism-like Mechanism Algorithm. 2014 IEEE International Conference on Computer and Information Technology, DOI 10.1109/CIT.2014.173.

Zarandi, M.H.F., Kazemi, A., 2008. Application of Rough Set Theory in Data Mining for Decision Support. Journal of Industrial Engineering, 1, 25–34.