

Fuzzy Rough Sets Theory Reducts for Quantitative Decisions – Approach for Spatial Data Generalization

Anna Fiedukowicz^(✉)

Faculty of Geodesy and Cartography, Department of Cartography,
Warsaw University of Technology, Warsaw, Poland
a.fiedukowicz@gik.pw.edu.pl

Abstract. One of the most important objectives within the scope of current cartography is the creation of system controlling the process of geographical data generalisation. Firstly, it requires selection of the features crucial from the point of view of the decision making process. Such tools as reducts and fuzzy reducts, though useful, are still insufficient for the quantitative decisions, common in cartographical generalization. Thus the author proposed a modification in fuzzy reducts calculating, which can allow to calculate them with regard to a continuous decision variable. The proposed method is based on the t-norm of fuzzy indiscernibility based on attribute value and fuzzy indiscernibility based on decision, which is calculated for each pair of objects. The solution seems to be more intuitive than the ones established previously.

Keywords: Fuzzy rough sets · Feature selection · Generalisation of geographic information · Cartography

1 Introduction

One of the most important challenges of modern cartography is the automation of the geographical information generalisation processes [1, 4, 9, 10]. It requires acquisition of data's crucial information, general patterns, and tendencies and their subsequent retaining on the lower levels of detail (LoD), which corresponds to lower map scales. This task up to now has been tackled manually by the skilled cartographers and it seems to be difficult to algorithmize. Even though there is an array of algorithms of generalisation, which address particular generalisation operators (such as: objects selection, simplification, smoothing, aggregation, amplification, etc.), what still poses the problem is the control of the entire process of generalisation – starting from the decision, “whether to generalise at all”, through the choice of the appropriate operator and algorithm, up to the final selection of parameters of the latter [1, 4, 9, 10].

The facts described above point to the conclusion that, apart from the particular tools employed in the process of generalisation, the decision-making system is required in order to manage the operations on many different levels. Such system should possess and utilise the skilled cartographer's knowledge. However, such skills usually result from years of practice and experience along with refined aesthetic taste, and therefore they are not available explicitly. Taking the above into consideration, according to the

author, the methods of knowledge data discovery (KDD) might be used to convert this hidden knowledge into an explicit form [5, 10].

Therefore the author's ultimate aim is to create a base of fuzzy rules governing the process of generalisation. The step prior to that, then, will be to choose the features crucial for the process of generalisation itself, which will lead to simplification of the decision system. The following paper addresses the problem of attribute choice methodology (using the reducts) which takes into consideration the specific data features. While in the previous work [10] the classical rough set approach for categorical data was used, this paper focuses on the problem of numeric feature selection based on numeric decisions – it discusses and provides some new extensions for fuzzy rough set (FRST) approach. The rough set approaches are chosen for the feature selection as they are easily understandable and give good intuition about why certain attribute is selected.

2 Data Specifics

The currently developed geospatial databases provide an array of information - in the form of attributes (projected in database structure), as well as more implicit features (connected with objects' geometry and their topography) - which can be used in the process of generalisation of geographical information. In that way the data is specified by a number of attributes that can potentially be used in the further generalisation process management.

What is worth emphasising, is the fact that the attributes can be expressed in different measurement scales: qualitative (ranging from binary scale, through classifying scale, to ordinal scale) as well as quantitative. Thus, the decision attribute can also be represented in different measurement scales.

In this article selection generalisation operator is considered, with decision expressed in two measurement scales (Table 1):

1. Binary – for the systems created on one particular data level of detail (corresponding with the scale 1:20 000): 1 - the object is selected, 0 - object is not selected;
2. Quantitative – for the system with universal character allowing to choose objects on any map scale (within assumed range) – the attribute's value is a corresponding scale denominator.

However, the second above is strongly preferred. Firstly it does not require designing separate systems for each of the desired scales. In the past, when analogue maps prevailed, it was possible to distinguish the scales in which the data were to be generated (they corresponded to the scales in which the maps were printed). However, nowadays most maps are accessible interactively via the Internet and the end user can choose any scale, thus the generalisation to all levels of detail is useful.

The test dataset corresponds to the topographical data collected at the 1:10 000 level of detail, which is available for the whole Poland's area in National Cartographical Database (pl. *Państwowy Zasób Kartograficzny*) and known as BDOT10 k (pl. *Baza Danych Obiektów Topograficznych* - Topographical Objects Database). However the data are strongly simplified (Fig. 1).

Table 1. Attributes values of test dataset (buildings): a_1 – building function (r – residential, o – office, s – shops & services, g – religious), a_2 – public function (1 – yes, 0 – no), a_3 – area (in square meters), a_4 – shortest distance to the river, a_5 – shortest distance to the road, a_6 – shortest distance to another building, a_7 – shortest distance to the forest, a_8 – shortest distance to built-up area; attributes a_3 to a_8 are calculated basing on objects geometry, a_4 to a_8 are expressed in meters; decision attributes (established by an expert) in different scales: dec_1 – quantitative scale, dec_2 – binary scale (for the LoD 1: 20 000 – $dec_2 = 1$ for $dec_1 \geq 20\,000$)

ID	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	dec_1	dec_2
1	r	0	101	304	35	43	126	0	0	15000
2	r	0	149	275	33	43	156	0	0	15000
3	r	0	93	252	37	41	200	0	0	15000
4	r	0	284	222	40	41	238	0	0	15000
8	r	0	892	216	124	76	212	0	0	17000
5	g	1	1721	32	53	420	537	0	1	50000
6	s	0	796	114	26	26	327	0	1	25000
7	s	0	585	122	26	26	321	0	1	25000
9	s	1	5174	70	94	143	234	0	1	40000
10	o	1	2840	69	54	95	140	0	1	35000
11	o	0	2015	19	19	46	394	0	1	27000
12	r	0	130	855	327	581	20	542	1	23000



Fig. 1. Graphical representation of test dataset (numbered buildings) with other objects assuring spatial context: forests, roads, river, built-up area

3 Rough Set Based Feature Selection

3.1 Rough Sets

The rough sets theory allows to reduce the complexity of a system by searching of reduct B – the subset of the entire attribute A set [6–8, 11]. The following search is based on the discernibility relation, which can be defined as:

$$R_B = \{(x, y) \in X^2 \text{ and } (\forall a \in B)(a(x) = a(y))\}$$

The so called decision reduct ensure the preservation of the original discernibility towards the decision: If the objects from different decision classes are discernible on the attribute set A, they are also discernible on its subset $B \in A$, being a reduct. The reduct has a minimal character, which means that none of the reduct’s attributes can be omitted without losing of discernibility mentioned above [6–8].

The approach described is connected with particular constraint: attributes as well as the decisions should be expressed in the classification (not orderly) scale. Other ways, a discretisation is required, what entails a partial loss of information (including e.g. the order of distinguished classes).

One of the extensions for this approach considers graded indiscernibility between objects. Thus, the classes of attributes can be more or less similar to each other [8]. Established dissimilarities between attribute classes – degrees of discernibility can be expressed in the form of a matrix (example – Table 2).

Table 2. Different discernibility degrees for classes of attribute a_1

a_1	r	s	o	g
r	0	0.8	1	1
s	0.8	0	0.5	1
o	1	0.5	0	1
g	1	1	1	0

3.2 Dominance-Based Rough Set Approach

The dominance based rough set (DBRS) approach, which is an extention of rough set theory, enables, without the loss of information, the use of attributes as well as decisions expressed in the ordinal scale. The theory postulates the approximation (and consequently reducts’ calculation) for the union of the subsequent decision classes. The theory is insufficient, however, in the case of the attributes expressed in quantitative scales, as it indeed assumes the monotonous relation between the attributes, but does not establish the distance between the subsequent classes [3].

3.3 Fuzzy Rough Sets

The hybrid of the fuzzy sets theory and rough sets theory, enabling to create fuzzy reducts, employs the attributes in quantitative scale. The discernibility relation based on

the equality of the attributes was replaced with the measure expressing the closeness of the objects represented by a fuzzy discernibility relation R [2, 11, 12].

The discernibility matrix then, existing also in the traditional fuzzy sets theory, is filled with the measure of closeness (based on each attribute) for each pair of objects with different decision value. In this paper the value of the fuzzy indiscernibility is calculated as follows:

$$R_a(x, y) = \frac{|a(x) - a(y)|}{l(a)} \quad (1)$$

Subsequently, on the basis of the discernibility matrix quality of the reduct can be calculated by finding best fuzzy discernibility R_b for each pair of objects and finding its' minimal value out of all pairs (instead of max any co-norm \perp can be used):

$$q(B) = \min(\max(R_{b \in B})) \quad (2)$$

This approach is based on the original RST assumption that each reduct is as good as its weakest component, meaning it is as good as the least discernible pair of objects belonging to separate decision classes. Therefore, in the original approach minimum operator is used, however some authors find it overly restrictive and allow the use of an average instead [2]:

$$q(B) = \text{mean}(\max(R_{b \in B})) \quad (3)$$

Such approach was used in the following work. The quality $q(B)$ can be then compared with the quality of the whole attribute set A (where ε is the acceptable tolerance of the quality loose) [2]:

$$q(B) \geq (1 - \varepsilon)q(A) \quad (4)$$

Another approaches for reduct evaluation is to punish pairs of objects (x, y) which belong to different decision classes but are nearly indiscernible using reduct's attributes. It can be expressed as [12]:

$$P_B = \mathcal{T}_{b \in B}(1 - R_b(x, y)) \quad (5)$$

The FRST approaches are the first enabling the calculation of the reducts for the data presented at the beginning, without discretization. It is possible, however, to establish only the reduct for the first variant (1) – binary decision scale.

4 Fuzzy Rough Reducts for Quantitative Decision

4.1 Adaptations of Existing FRST Methods

Few among the authors directly address the problem of reducts for the decisions in the quantitative scale (dec_2), however some of the solutions from Sect. 3.3 can be adapted for this case.

In the formula (2) or (3) similarity of objects by decision can be added (now all the objects' pair are compared):

$$q(B) = \text{mean}(\max(R_{b \in B}, 1 - R_{dec})) \tag{6}$$

However, this solution disadvantage is that it promotes too much the pairs of objects which are indiscernible according to the decision ($R_{dec} \cong 0$) which is in fact not interesting when looking for decisive reducts.

Also formula (5) can be adapted by adding fuzzy indiscernibility relation to the t-norm [12]:

$$\mathcal{T}(P_B, R_{dec}) \tag{7}$$

The final punishment is calculated as a sum:

$$Sim(d/B) = \sum_{x,y:d(x) \neq d(y)} R_{dec} \prod_{b \in B} (1 - R_b) \tag{8}$$

The disadvantage of this approach (formulas 7 and 8) is that it does not seem intuitive as it includes indiscernibility by decision for data already aggregated by t-norm.

4.2 Proposed Solutions

The proposed solution intends to be more intuitive for non-mathematical expert. It is based on the necessity to establish the value of the relative relation R_{a_rel} considering the objects' relation R both on attribute (R_a) and decision (R_{dec}). Therefore, it is proposed to calculate the relative tolerance relation for each pair of objects as a t-norm of R_a and R_{dec} :

$$R_{a_rel} = \mathcal{T}(R_a, R_{dec}) \tag{9}$$

The most interesting from the point of view of applications described in introduction, seem to be such t-norms as:

1. product(R_a, R_{dec})
2. Hamacher product(R_a, R_{dec}): $\mathcal{T}_{H_0}(a, b) = \begin{cases} 0 & \text{if } a = b = 0 \\ \frac{ab}{a+b-ab} & \text{otherwise} \end{cases}$
3. min(R_a, R_{dec})

The further proceedings are identical as in the classic FRST method, though all possible pairs of objects are compiled and R_{b_rel} is used instead of R_b :

$$q(B) = \text{mean}(\max(R_{a_rel (b \in B)})) \quad (10)$$

Such an approach allows to follow the significance of each attribute (in relation to the decision) for each pair of objects. This can be valuable from the point of view of expert using decision system as it allows to intuitively understand the importance of attributes.

5 Experiments on Test Data

5.1 Some General Assumptions

The calculation of fuzzy discernibility indicator R_a for the pair of objects depends on the scale in which the attribute a was depicted. Therefore:

- For the attribute a_2 (expressed in binary scale) the classical discernibility approach, based on the equivalence relation was employed,
- For the attribute a_1 (expressed in the classifying scale) similarity matrix was employed (Table 2),
- For the attributes a_3 to a_8 (in quantitative scale) tolerance relation R basing on formula (1) was employed.

Due to the specificity of the described problem establishing of all possible reducts of the set was not necessary. In practice, for the purpose of the further application only 1, sometimes few, reducts will be used. The accessibility to the attributes necessary for the calculation is usually high as they are available in the databases either as the descriptive attributes designed in database structure, or are easy to calculate on the basis of objects geometry. Therefore, the reducts were calculated with Johannson's heuristic. It operates as follows: every time such attribute is added to the reduct, which results in the biggest increase of the quality q (understood as in formula 3). This steps may be repeated until:

- Obtaining the quality q fulfils the condition (4) assumed by the user or
- The point when adding another attributes results in the increase of quality q lower then established Δq .

In this work the second method (with $\Delta q = 0.02$) was employed, due to the necessity of maintaining a low system complexity (and consequently not overly numerous reducts), if its higher complexity did not increase the overall quality significantly.

Similarly, the other reducts can be calculated (starting from the subsequent attribute), however this work limited itself to only 1 reduct in each example.

Table 3. Subsequent steps of fuzzy reduct creation for the binary decision, including the corresponding reduct qualities (the fields with the highest accuracy are highlighted, while the elements added to the reduct are in bold)

No	current reduct	quality	quality of a reduct after adding new attribute:							
			a ₁	a ₂	a ₃	a ₄	a ₅	a ₆	a ₇	a ₈
1	{}	0,00	0,77	0,43	0,33	0,29	0,22	0,28	0,30	0,14
2	{a ₁ }	0,77	0,77	0,80	0,80	0,87	0,90	0,91	0,82	0,91
3	{a ₁ , a ₆ }	0,91	0,91	0,94	0,93	0,91	0,91	0,91	0,91	0,91
4	{a ₁ , a ₂ , a ₆ }	0,94	0,94	0,94	0,94	0,94	0,94	0,94	0,94	0,94
5	{a₁, a₂, a₆}	0,94								

5.2 Fuzzy Reducts for Binary Decision

Determination of the reduct started with combining the objects belonging to different decision classes (*dec_i*) into pairs and calculating the discernibility matrix of relations R (according to the rules described in Sect. 5.1). Then the consecutive elements of the reduct were established with use of greedy heuristic basing on the *q* value – Table 3.

Consequently reduct {a₁, a₂, a₆} have been established. Adding another attribute would not increase the quality of the reduct so no other attribute was added. What is worth mentioning is the high quality of the decision reduct, as the ability to distinguish the object with different decisions, is high.

Table 4. Subsequent steps of fuzzy reduct creation for the quantitative decision, including the corresponding qualities (highlights as in Table 3); the method used: according to the formula 6

No	current reduct	quality	quality of a reduct after adding new attribute:							
			a ₁	a ₂	a ₃	a ₄	a ₅	a ₆	a ₇	a ₈
1	{}	0,00	0,90	0,90	0,73	0,68	0,67	0,73	0,71	0,68
2	{a ₁ }	0,90	0,90	0,94	0,92	0,90	0,91	0,91	0,90	0,92
3	{a ₁ , a ₂ }	0,94	0,94	0,94	0,94	0,94	0,96	0,96	0,94	0,96
4	{a ₁ , a ₂ , a ₆ }	0,96	0,96	0,96	0,96	0,96	0,96	0,96	0,96	0,96
5	{a₁, a₂, a₆}	0,96	or {a ₁ , a ₂ , a ₅ } or {a ₁ , a ₂ , a ₈ }							

5.3 Fuzzy Reducts for Quantitative Decisions

First the method using formula 6 was used. Results of following steps are presented in Table 4. One of the possible decision reduct is exactly the same that the one described in Sect. 5.2, although it allows to distinguish a number of object types. However accuracies seem to be over-optimistic since the goal was to distinguish between more exact decisions and the qualities here are higher than for the binary decision.

The next step was to test the method proposed in Sect. 4.2. In this case all possible pairs of objects were juxtaposed. For each of them *R_α*, *R_{dec1}* and *R_{α-rel}* were calculated (in two variants: with use of T-norm product and Hamacher product). Then the quality

Table 5. Subsequent steps of fuzzy reduct creation for the quantitative decision, including the corresponding qualities (highlights as in Table 3); the t-norm used: $R_{rel} = \text{product}(R_a, R_{dec})$

No	current reduct	quality	quality of a reduct after adding new attribute:							
			a ₁	a ₂	a ₃	a ₄	a ₅	a ₆	a ₇	a ₈
1	{}	0,00	0,30	0,26	0,15	0,10	0,07	0,13	0,13	0,04
2	{a ₁ }	0,30	0,30	0,33	0,32	0,31	0,32	0,32	0,31	0,31
3	{a ₁ , a ₂ }	0,32	0,33	0,33	0,33	0,34	0,35	0,35	0,34	0,35
4	{a ₁ , a ₂ , a ₆ }	0,35	0,35	0,35	0,35	0,35	0,35	0,35	0,35	0,35
5	{a₁, a₂, a₆}	0,35	or {a ₁ , a ₂ , a ₅ } or {a ₁ , a ₂ , a ₈ }							

Table 6. Subsequent steps of fuzzy reduct creation for the quantitative decision, including the corresponding qualities (highlights as in Table 3); the t-norm used: $R_{rel} = \text{Hamacher product}(R_a, R_{dec})$

No	current reduct	quality	quality of a reduct after adding new attribute:							
			a ₁	a ₂	a ₃	a ₄	a ₅	a ₆	a ₇	a ₈
1	{}	0,00	0,31	0,26	0,19	0,13	0,10	0,16	0,18	0,04
2	{a ₁ }	0,31	0,31	0,34	0,34	0,33	0,34	0,33	0,33	0,33
3	{a ₁ , a ₂ }	0,34	0,33	0,33	0,33	0,34	0,35	0,35	0,34	0,35
4	{a ₁ , a ₂ , a ₆ }	0,35	0,36	0,36	0,36	0,36	0,36	0,36	0,36	0,36
5	{a₁, a₂, a₆}	0,35	or {a ₁ , a ₂ , a ₅ } or {a ₁ , a ₂ , a ₈ }							

Table 7. Subsequent steps of fuzzy reduct creation for the quantitative decision (decision artificially brought to the binary scale), including the corresponding accuracies (highlights as in Table 3); the t-norm used: product and Hamacher product (results for both are identical)

No	current reduct	quality	quality of a reduct after adding new attribute:							
			a ₁	a ₂	a ₃	a ₄	a ₅	a ₆	a ₇	a ₈
1	{}	0,00	0,41	0,23	0,17	0,15	0,12	0,15	0,16	0,08
2	{a ₁ }	0,41	0,41	0,42	0,42	0,46	0,48	0,48	0,43	0,48
3	{a ₁ , a ₆ }	0,48	0,48	0,50	0,49	0,48	0,48	0,48	0,48	0,48
4	{a ₁ , a ₂ , a ₆ }	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50	0,50
5	{a₁, a₂, a₆}	0,50								

based on R_{a_rel} was calculated for particular attributes – the subsequent steps of the heuristics are illustrated in the tables below (Tables 5 and 6).

The methods results in the same reducts as the previous ones, irrespectively which of the t-norm is used. What is more, even though qualities values differ depending on the used t-norm (and differ even more form the corresponding ones in Table 4), there seem to be noticeable tendencies in its distribution over the attributes.

In the next stage, in order to test the universality of the method, the same steps were taken, though in this case on binary decision (dec_1) – Table 7. In this case the results, irrespectively of the chosen t-norm, were alike ($R_{dec} \in \{0, 1\}$, so all t-norms have the same value in the R_b function). It should be noted, that the result (calculated reduct) and

the relation between the qualities of particular reducts are identical to those calculated with the classic method described in Sect. 5.2. The ratio of corresponding accuracies in the Tables 3 and 7 equals c.a. 0.53, which reflects the proportion of the number of pairs in different decision classes and the total number of pairs (35/66), or in other words the average discernibility of all pair of objects towards the decision.

6 Conclusions

The work above addressed the problem of reduct calculation in the case of decisions in quantitative scale. In the process a tolerance relation (R_{rel}), understood as t-norm of tolerance relations of attribute (R_a) and decision (R_{dec}), was employed.

The conducted tests using the two types of t-norm – product and Hamacher product – gave similar results, meaning they both allowed to achieve the same reducts irrespective of differences in the quality value of the reduct. Generally the reducts' qualities values calculated with this method were lower than for binary decision, what seems justifiable taking into consideration the necessity to distinguish the objects according to their continuous decision value. They are also lower than the qualities calculated by one of the existing methods (formula 6), however those qualities seem to be over-optimistic as they grow unreasonably thanks to the objects having the same or similar decision values.

The methodology employed in the case of artificially binary decision allowed to achieve the same reduct as it was the case in the original FRST method. However, even though the relation of quality values between particular attributes were maintained the absolute values of quality were different. It is a result of calculating the discernibility relation for all pair of objects.

The tests indicate that proposed method can be applied in the process of generalisation of geographical data mentioned above. However there is also a potential for other application areas. Depending on the application type other t-norm for calculating R_{rel} can be used.

The main advantage over the other method is that it is intuitively understandable and prevents black-box solution, allowing user to follow the importance of each attribute on every single step of the reduct computation. Therefore the method can be employed among others in the creation of the system of generalisation control, on the first stage of its development – attribute selection.

Acknowledgments. Author would like to thank Prof. Dominik Ślęzak for consultations and benevolent intellectual support.

References

1. Burghardt, D., Duchene, C., Mackaness, W. (eds.): Abstracting Geographic Information in a Data Rich World. Lecture Notes in Geoinformation and Cartography Series. Springer, Berlin (2014)
2. Cornelis, Ch., Jensen, R., Martín, G.H., Slezak, D.: Attribute selection with fuzzy decision reducts. Inf. Sci. **180**(2), 209–224 (2010)

3. Greco, S., Matarazzo, B., Słowiński, R.: Multicriteria classification by dominance-based rough set approach. In: Kloesgen, W., Zytkow, J. (eds.) *Handbook of Data Mining and Knowledge Discovery*. Oxford University Press, New York (2002)
4. Mackaness W.: Understanding geographic space. In: Mackaness, w., Ruas, A., Sarjakoski, T. (eds.) *Generalisation of Geographic Information: Cartographic Modelling and Application*. Elsevier, Oxford (2007)
5. Miller, H.J., Han, J.: *Geographic Data Mining and Knowledge Discovery*. Taylor & Francis, London (2001)
6. Pawlak, Z.: Rough sets. *Int. J. Parallel Prog.* **11**(5), 341–356 (1982)
7. Pawlak, Z.: *Rough Sets: Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishing, Dordrecht (1991)
8. Pawlak, Z., Skowron, A.: Rough sets: some extensions. *Inf. Sci.* **177**, 28–40 (2007)
9. Olszewski R., Kartograficzne modelowanie rzeźby terenu metodami inteligencji obliczeniowej, *Prace Naukowe - Geodezja*, z. 46, Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa (2009)
10. Olszewski, R., Fiedukowicz, A.: Supporting the process of monument classification based on reducts, decision rules and neural networks. In: Kryszkiewicz, M., Cornelis, C., Ciucci, D., Medina-Moreno, J., Motoda, H., Raś, Z.W. (eds.) *RSEISP 2014. LNCS*, vol. 8537, pp. 327–334. Springer, Heidelberg (2014)
11. Riza, L.S., Janusz, A., Bergmeir, C., Cornelis, C., Herrera, F., Ślęzak, D., Benítez, J.M.: Implementing algorithms of rough set theory and fuzzy rough set theory in the R package “RoughSets”. *Inf. Sci.* **287**, 68–89 (2014)
12. Ślęzak, D., Betliński, P.: A role of (not) crisp discernibility in rough set approach to numeric feature selection. In: Hassanien, A.E., Kim, T.-H., Ramadan, R., Salem, A.-B.M. (eds.) *AMLTA 2012. CCIS*, vol. 322, pp. 13–23. Springer, Heidelberg (2012)