RESEARCH OF BEHAVIOUR OF FUZZY CORA ON SELECTIONS OF DIFFERENT VOLUMES

Roman Grekov and Arkady Borisov

Keywords: pattern recognition; recognition threshold; fuzzy relations; fuzzy objects

Introduction

The main idea of the proposed method is a consistent division of a description space into lower dimensions description spaces. The result of this division is two-dimensional fuzzy set shadows, represented as binary relations. By performing compositions of near by relations is possible to get degrees of reductions, which characterize how powerful the connection between these relations is. These degrees of reduction can be converted such as to give evaluations to every feature participated in recognition and/or to combinations of features. Next, after the most effective of them are selected (i.e. with a maximal evaluation), a decision rule is created. With a participation of this rule the recognition is performed [3]. The variation of the method with participations of Cross-Validation [4] offers a partitioning of the initial set of objects into a finite number of subsets and to perform a learning procedure and to test recognition on each of these subsets separately. After an error has been calculated for each subset, the recognition of objects is performed with participations of the decision rule derived during learning and test recognitions on subset with a minimal error (i.e. with a maximal number of successfully recognized test objects).

Initialisation of experiments

To check the performance of offered method Fuzzy Cora [5] a finite number of experiments need to be performed. On this stage a variant of examination of the method by performing a feature evaluation and recognition for tasks of different structure and size is introduced. To choose a test set correctly the major objective of implementations of experiments needs to be defined.

1. Evaluate an efficiency of the method.
2. Evaluate an efficiency of the method on a high volume task sets while changing a volume of a learning set.
3. Compare the work of offered method with participation of Cross-Validation to the method of recognition with participation of compositions of binary relations.

To start lets define maximal and minimal space dimension to perform experiments on. Primarily, it is limited by hardware resources, which mainly needed for the process of learning. Table 1 represents the time needed for the process of learning for objects of tasks of different sizes (when dividing an initial set of objects with Cross-Validation into three subsets), using specially developed software:

Table 1. Time, consumed to perform different tasks

Type of the task		Time				
Dimension (the number of features)	Values of the feature (the number of values feature can have)	Time, consumed to prepare task space	Time, consumed to pre-fill the space of features	Time, consumed to create shadows of fuzzy sets	Time, consumed for recognition	Total Time
3	5	4	2	6	1	13
3	7	5	1	19	1	26
3	10	16	11	1:43	3	2:13
4	3	8	1	13	4	26
4	5	18	5	2.48	4	3:15
4	6	33	9	7.56	3	8:41
4	10	4:31	2.54	4:16:39	9	4:24:13
5	3	30	2	1:46	12	2:30
5	5	2:01	32	1:35:43	13	1:38:29

Thus, because of limitations in time and hardware let us define maximal object description spaces. Lets use following task examples (series):

1) 3-dimensional object space with the number of values a feature can have up to 10.
2) 4-dimensional object space with the number of values a feature can have up to 7.
3) 5-dimensional object space with the number of values a feature can have up to 5.

Next, define centres of object classes, using an exponential function (example of the function shown is for the 4-dimensional object space):

$$\mu_A = \exp[-((X_1 - K_1)^2 + (X_2 - K_2)^2 + (X_3 - K_3)^2 + (X_4 - K_4)^2)]$$

$$\mu_B = \exp[-((X_1 - L_1)^2 + (X_2 - L_2)^2 + (X_3 - L_3)^2 + (X_4 - L_4)^2)]$$

where

μ_A and μ_B - membership functions to classes *A* and *B*;

K_1, K_2, K_3, K_4 - variables, which characterize coordinates of the centre of 4-dimensional ellipse of the class *A*;

L_1, L_2, L_3, L_4 - variables, which characterize coordinates of the centre of 4-dimensional ellipse of the class *B*.

Also, each series includes spaces of features of the same dimension. Features will vary by the number of possible values feature can have. The scheme of experiments is shown in Figure 2.

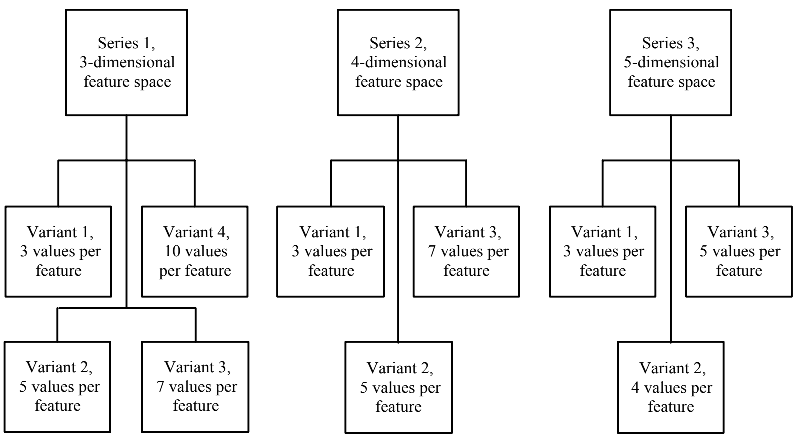

Fig. 2. Structural division of experiments

Hence, each variant is a set of experiments, which differs only by the position of classes in an object space. The choice of the centre of each class is very important. The centre of a class defines a disposition of objects in the space. The closer object is to the class the higher value of the membership function to that class this object has. Keeping in mind these ideas it is possible to choose such object description spaces, which definitely show the partition of objects in the space. I. e. the farther centres of different classes stand to each other, the greater a distance between these classes is, and vice versa. Also, putting centres of classes in one plane correspondingly to each other, it is possible to get the situation, where only one feature is a definitive one. In this situation centres of classes should differ only by one index. For instance, in 3-dimensional case the centre of class A could be in (2, 2, 1) and the centre of class B – in (2, 2, 3). For the situation where the proposed method needs to be checked in a badly partition able task, coordinates of centres of classes should stand closer to each other. This brings up a minimal partitioning between classes. For this particular task lets choose such an object set, as centres of classes were positioned in the same way for different series. For example, in 4-dimensional space these will be sets of objects with centres of classes in most distant to each other areas like (1, 1, 1, 1) and (3, 3, 3, 3), with a condition that maximum value features can have is 3. Analogically, for the example in 3-dimensional object description space (with the same conditions) these will be centres in (1, 1, 1) and (3, 3, 3).

Now define volumes of learning and test sets. Lets put a percentage as a limit (software developed is able to randomise a choice of elements using a defined percentage) – learning set is 40, 60 and 80% of the maximal quantity of objects in the space for different experiments and test set is the rest of 60, 40 and 20% correspondingly. In the case of Cross-Validation the percentage is using to divide initial object space to a learning set and a set with unknown values of a membership function. Next, learning set is dividing by the degree of Cross-Validation (2) and then every set we get is dividing onto learning and test sets again using known percentage (in our case it is 30%). For instance, in case of 3x3 the next volumes of sets will appear:

Total number of objects in space	27
Recognition set	11
Common learning set	16
Test set CV1	2
Test set CV2	2
Learning set CV1	6
Learning set CV2	6

Hence, correspondingly to each series the table 2 showing the number of objects in every set is defined:

Table 2. Description Sets for proposed Experiments

Number of Series	Type of the space	Total number of objects in the space	Number of objects in sets	
			Learning set (objects)	Test set (objects)
1	3 x 3 x 3	27	22	5
2	5 x 5 x 5	125	100	25
3	7 x 7 x 7	343	257	86
4	10 x 10 x 10	1000	800	200
5	3 x 3 x 3 x 3	81	65	16
6	5 x 5 x 5 x 5	625	500	125
7	7 x 7 x 7 x 7	2401	1921	480
8	3 x 3 x 3 x 3 x 3	243	194	49
9	4 x 4 x 4 x 4 x 4	1024	819	205
10	5 x 5 x 5 x 5 x 5	3125	2500	625

Next stage is a run of experiments and memorizing of experimental results for further comparison.

Run of Experiments

Next, variants of recognitions for chosen object description spaces and examples of tasks are presented [3]. The selection of a recognition threshold is chosen individually for each experiment. Threshold is the value needed to compare feature efficiency with. Only features with the value greater than threshold are selected. For current experiments variants without a participation of Cross-Validation and using Cross-Validation (N/A cases) with the value of degree as 2 were selected.

Table 3. Results of experiments

Number of experim ent	Dimensi on of the space	Number of values per feature	Degree of Cross-Validati on	Number of objects in the space	Number of objects in learning set, %	Number of objects in learning set	Number of objects in recognition set	Number of successfully recognized objects	Successfu lly recognize d obects, %
1	3	3	N/A	27	40	11	16	10	62,50
2	3	3	N/A	27	60	16	11	7	63,64
3	3	3	N/A	27	80	22	5	5	100,00
4	3	3	2	27	40	11	16	5	31,25
5	3	3	2	27	60	16	11	5	45,45
6	3	3	2	27	80	22	5	3	60,00
7	3	5	N/A	125	40	50	75	50	66,67
8	3	5	N/A	125	60	75	50	40	80,00
9	3	5	N/A	125	80	100	25	19	76,00
10	3	5	2	125	40	50	75	42	56,00
11	3	5	2	125	60	75	50	33	66,00
12	3	5	2	125	80	100	25	17	68,00
13	3	7	N/A	343	40	137	206	108	52,43
14	3	7	N/A	343	60	206	137	70	51,09
15	3	7	N/A	343	80	274	69	41	59,42
16	3	7	2	343	40	137	206	110	53,40
17	3	7	2	343	60	206	137	73	53,28
18	3	7	2	343	80	274	69	40	57,97
19	3	10	N/A	1000	40	400	600	314	52,33
20	3	10	N/A	1000	60	600	400	142	35,50
21	3	10	N/A	1000	80	800	200	59	29,50
22	3	10	2	1000	40	400	600	80	13,33
23	3	10	2	1000	60	600	400	191	47,75
24	3	10	2	1000	80	800	200	71	35,50
25	4	3	N/A	81	40	32	49	35	71,43
26	4	3	N/A	81	60	49	32	24	75,00
27	4	3	N/A	81	80	65	16	10	62,50
28	4	3	2	81	40	32	49	31	63,27
29	4	3	2	81	60	49	32	27	84,38
30	4	3	2	81	80	65	16	14	87,50
31	4	5	N/A	625	40	250	375	298	79,47
32	4	5	N/A	625	60	375	250	204	81,60
33	4	5	N/A	625	80	500	125	108	86,40
34	4	5	2	625	40	250	375	275	73,33
35	4	5	2	625	60	375	250	200	80,00
36	4	5	2	625	80	500	125	101	80,80
37	4	7	N/A	2401	40	960	1441	379	26,30
38	4	7	N/A	2401	60	1441	960	279	29,06
39	4	7	N/A	2401	80	1921	480	146	30,42
40	4	7	2	2401	40	960	1441	396	27,48
41	4	7	2	2401	60	1441	960	251	26,15
42	4	7	2	2401	80	1921	480	119	24,79
43	5	3	N/A	243	40	97	146	101	69,18
44	5	3	N/A	243	60	146	97	78	80,41
45	5	3	N/A	243	80	194	49	37	75,51
46	5	3	2	243	40	97	146	100	68,49

Table 3. Results of experiments (cont.)

47	5	3	2	243	60	146	97	71	73,20
48	5	3	2	243	80	194	49	36	73,47
49	5	4	N/A	1024	40	410	614	543	88,44
50	5	4	N/A	1024	60	614	410	383	93,41
51	5	4	N/A	1024	80	819	205	195	95,12
52	5	4	2	1024	40	410	614	546	88,93
53	5	4	2	1024	60	614	410	382	93,17
54	5	4	2	1024	80	819	205	194	94,63
55	5	5	N/A	3125	40	1250	1875	1188	63,36
56	5	5	N/A	3125	60	1875	1250	800	64,00
57	5	5	N/A	3125	80	2500	625	418	66,88
58	5	5	2	3125	40	1250	1875	1177	62,77
59	5	5	2	3125	60	1875	1250	813	65,04
60	5	5	2	3125	80	2500	625	425	68,00

In the next stage compare results of experiments to check the work of the method for different variants by following criteria:

a) experiments in spaces with the same dimension size (with the same number of features, but different values of the feature).
b) experiments in spaces with the same dimension size (with the same number of features and the same values of the feature). Compare results for variations of the method with/without Cross-Validation.
c) Experiments in spaces with different dimension sizes. Evaluate Cross-Validation in common when changing an object space.

For better understanding of the task lets represent received results with the help of the graphics. X-axis shows a percentage of learning set objects to the total number of objects in object description set and Y-axis is a percentage of successfully recognized objects. It's logical to split received results into groups by the number of features in a task. As the result three figures for tasks with participation of three, four and five features correspondingly were derived. Hence, by making an analysis of behaviour of recognition graphics it is possible to get a greater imagination of how good the method works and ways of to make it work better.

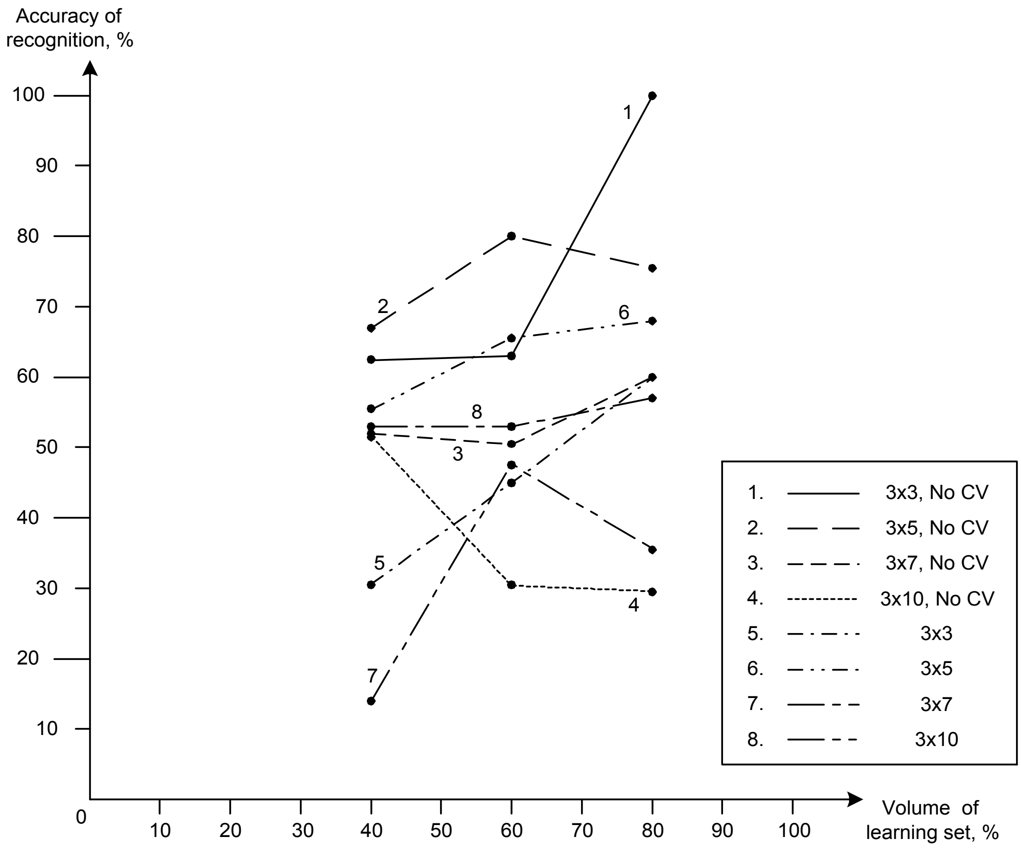

Рис. 4. Распознавание в 3-х мерном пространстве

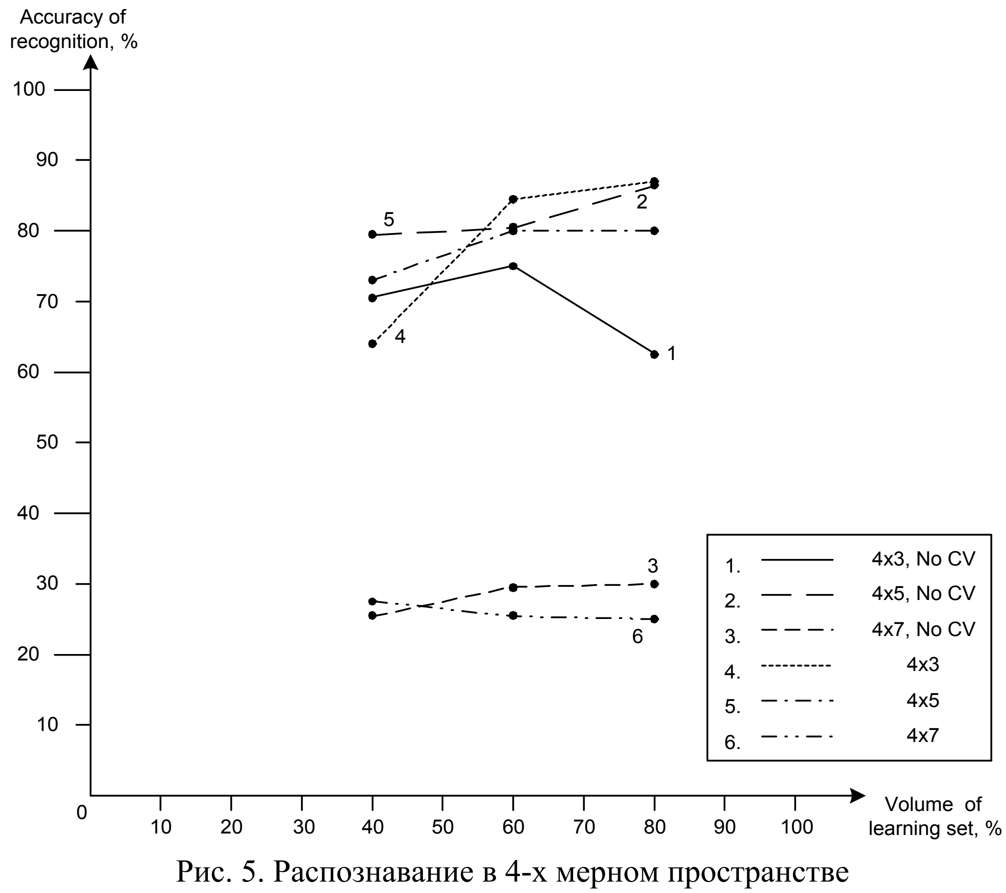

Рис. 5. Распознавание в 4-х мерном пространстве

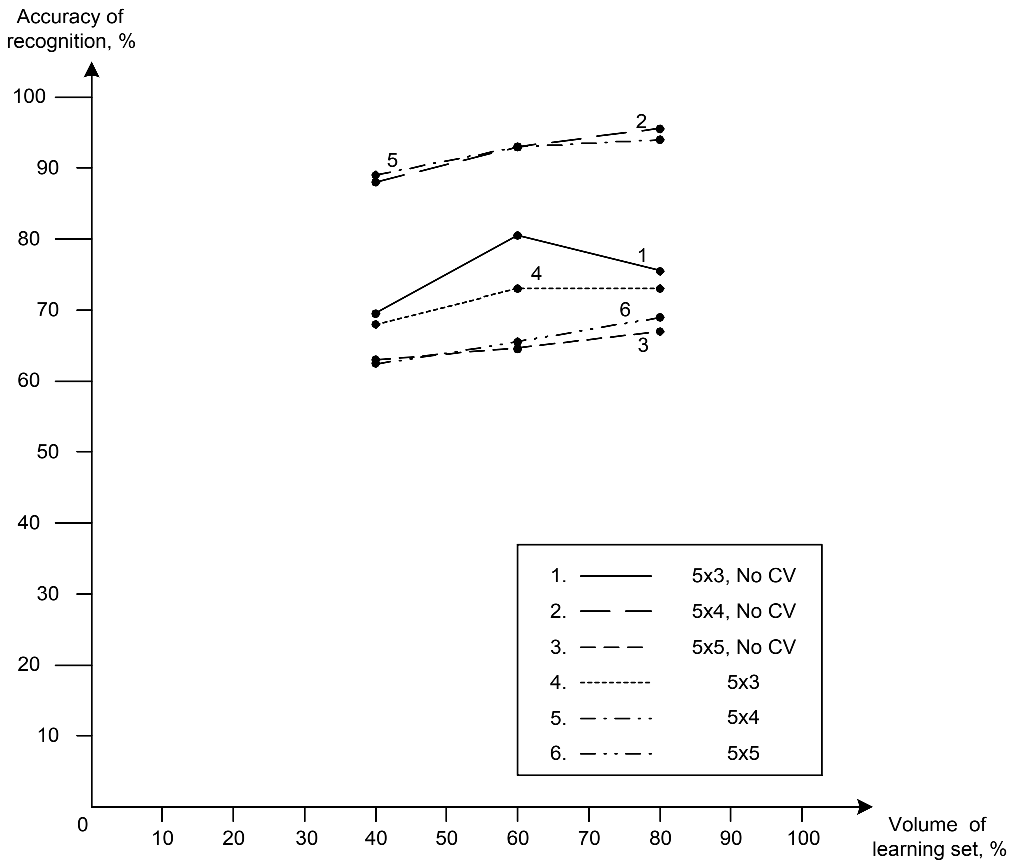

Рис. 6. Распознавание в 5-ти мерном пространстве

Using the graphics following conclusions were maid:

1) While increasing the volume of learning set the quality of recognition is improving insignificantly and is dependant on total number of objects in initial object space. For example, for the smallest object description space, on which the experiments were performed (3x3), when increasing a volume of training set the quality of recognition has increased by 30-35%. For cases, when the volume of an initial set is big enough, the improvement of quality of recognition is insignificant. For example, in case of 4x7. Thus, it is possible using experiments it is possible to determine the sufficient volume of a training set for each certain task. For a case 3x3, for example, it will be 75-80 % from the total number of objects of a task.
2) Using of a method of verification of qualifiers Cross-Validation in some cases has not increased an accuracy of recognition, as it was expected. It can be explained by some objective reasons, such, as, for example, a random choice of a training set. Thus, an idea to modify a method such as to perform test recognition with a method with participation of Cross-Validation, and also using all space (as it is done in a variant of a method without participation Cross-Validation), and after the comparison of results of test recognition select the best variant of a decision rule.
3) For some variants, like 4x7, accuracy of recognition is quite low. As it is visible from the graphics, it is specific when increasing the number of values feature can get. It's rather stipulated by using an exponential function to define elements of a description space and when a distance between classes is increasing a large number of objects

with a very weak membership to any of classes is appearing, i.e. membership function for these objects is aspiring to zero.

4) In case of 3-dimensional space of features using of Cross-Validation didn't bring expected results. The main reason is that the set of objects is quite low. As the main idea of Cross-Validation is in division of initial set of objects into the certain number of subsets, that when partitioning an initial set of a small power there is a known degree of uncertainty, which negatively influences recognition.

Hence, lets formulate basic theses, typical for the proposed method Fuzzy Cora, which is based on a combination of Cross-Validation and the method with participation of binary compositions:

1) In case of increasing the volume of learning set quality of recognition is improving, but in case when the number of objects and a volume of a learning set are increasing the quality of recognition is improving insignificantly.
2) Using of Cross-Validation in Fuzzy Cora algorithm is admissible and desirable while working with complicated tasks and in case of the high volume of an object set.
3) Performance of the method in high dimensional object spaces is quite good – in most cases the number of successfully recognised objects is greater than 60% and in some cases is not lower than 80%.

Literature

1. Zadeh, L. A. Fuzzy Sets and Systems. *Proceedings of the Symposium on Systems Theory*, Polytechnic Institute of Brooklyn, 1965.
2. Melikhov A.N., Bernshtein L.S., and Korovin S.Y. *Situacionnie sovetujuschie sistemi s nechetkoi logikoi*. Moscow, Nauka Publishers, 1990 (In Russian).
3. Borisov A., Ozols J. & Grekov R. *Construction of features and decision rules in fuzzy pattern recognition task*. Lectures Notes of the Nordic-Baltic Summer School on Intelligent Design, Intelligent Manufacturing and Intelligent Management, 1999.
4. Weiss S.M. & Kapouleas I. (1989). An empirical comparison of pattern recognition, neural nets and machine learning classification methods. *Proceedings of the 11-th International Joint Conference on Artificial Intelligence*. San Mateo, CA: Morgan Kaufmann.
5. Ozols, Y. and Borisov A. Pattern Classification and Feature Extraction on the Basis of Composition of Fuzzy Relations. Proceedings of the Seventh International Fuzzy Systems Association World Congress, IFSA'97, Prague, Czech Republic, June 25-29, 1997.

Romans Grekovs is a Ph.D. student in the Institute of Information Technology at the Technical University of Riga (Latvia). He has a M.Sc. from the Technical University of Riga. His research interests include fuzzy sets theory and fuzzy pattern recognition. Address: Decision Support Systems Group, Institute of Information Technology, Technical University of Riga, 1 Kalkyu Street, Riga LV-1658, Latvia. E-mail: roman_grekov@yahoo.com.

Grekovs R. Research of Behaviour of Fuzzy Cora on Selections of Different Volumes
This paper is dedicated to efficiency evaluations of a method Fuzzy Cora, which is based on an analysis of the structure of an initial space of pattern descriptions. This work reviews the behaviour of the method during the changing of a volume of a learning set using chosen parameters of the task.

Греков Р. Исследование Поведения Fuzzy Cora на Выборках Различного Объема
Данная работа посвящена оценке эффективности метода Fuzzy Cora, который основан на анализе структуры пространства описаний образов. Статья рассматривает поведение метода с изменением объема обучающего множества используя выбранные параметры задачи.

Grekovs R. Fuzzy Cora Metodes Pētīšana Dažāda Apjoma Izlasēs
Šis raksts veltīts Fuzzy Cora metodes efektivitātes novērtēšanai, kura ir balstīta uz tēlu apraksta telpas struktūras analīzes. Darbs arī apskata metodes uzvedību mainījot apmācošas kopas apjomu un izlietojot izvēlētus uzdevuma parametrus.