MULTI-VARIANT PYRAMIDAL CLUSTERING AND ANALYSIS HIGH-DIMENSIONAL DATA

Krassimira B. Ivanova, Vitalii Velychko, Krassimir Markov

Abstract: *In this work an example of multi-variant clustering is presented. The problems to be solved are described and multi-variant clustering based on pyramidal multi-layer multi-dimensional structures is outlined. The conclusion is that the multi-variant clustering combined with pyramidal generalization and pruning gives reliable results.*

Keywords: *Data mining, multi-variant clustering, pyramidal multi-layer multi-dimensional structures.*

ACM Classification Keywords: *H.2.8 Database Applications, Data mining; I.5.3 Clustering.*

Introduction

Clustering is a fundamental problem that has numerous applications in many disciplines. Clustering techniques are used to discover natural groups in data sets and to identify abstract structures that might reside there without having any background knowledge of the characteristics of the data. They have been used in a variety of areas, including bioinformatics; computer vision; VLSI design; data mining; gene expression analysis; image segmentation; information retrieval; information theory; machine learning; object, character, and pattern recognition; signal compression; text mining; and Web page clustering [Kogan, 2007].

Clustering systems build a generalization hierarchy by partitioning the set of examples in such a way that similarity is maximized within a partition and minimized between them. At the lowest level of the hierarchy are the individual examples.

Clustering is especially suited to unsupervised learning, where the concepts to be learned are not known in advance, but it may also be applied to learning from examples. A new example is classified by considering adding it to each cluster, and determining which one it fits best. This process is repeated down the hierarchy until a cluster is reached that contains only examples of a single class. The new example adopts the class of this cluster. The main differences between different clustering methods are the similarity measure, and the method used to evaluate each cluster to determine the best fit for the new example. Approaches range from Euclidean distance to Bayesian statistics. Clustering is therefore the broad approach of concept formation by grouping similar examples. [Luo et al, 2009]

Clustering has attracted research attention for more than 50 years. A partial list of excellent publications on the subject is provided in [Kogan, 2007].

Below we present a simple example of multi-variant clustering and analysis high-dimensional data based on multi-dimensional pyramidal multi-layer structures.

Similar structures were proposed by Prof. V.P. Gladun in a series of publications [Gladun, 1987, 1994, 2000; Gladun and Vashchenko, 2000, Gladun et al, 2008]. They were named as "*growing pyramidal networks*" (GPN).

Pyramidal network is a network memory, automatically tuned into the structure of incoming information. Unlike the neuron networks, the adaptation effect is attained without introduction of a priori network excess. Pyramidal networks are convenient for performing different operations of associative search. Hierarchical structure of the networks, which allows them to reflect the structure of composing objects and natural gender-species’ bonds, is an important property of pyramidal networks.

The further text of the paper is organized as follow. Firstly we describe the problems to be solved. In the next chapters we present an example of sparse high dimensional vectors and multi-variant clustering based on pyramidal multi-layer multi-dimensional structures. Ranging the variants using Dunn indices is given. Finally, the conclusions are outlined.

Basic problems to be solved

For a given set of instances $\mathbf{R} = \{R^i, i \in 1,...,r\}$ and a query Q one often is concerned with the following basic problems:

1. Find instances in $\mathbf{R}$ “related” to the query Q if, for example, a “distance” between two instances R_i and R_j is given by the function $d(R_i, R_j)$ and a threshold $tol > 0$ is specified one may be interested in identifying the subset of instances $\mathbf{R}_{tol} \subseteq \mathbf{R}$ defined by $\mathbf{R}_{tol} = \{R : R \in \mathbf{R}, d(Q,R) < tol\}$.
2. Partition the set $\mathbf{R}$ into disjoint sub-collections $\pi_1, \pi_1, ..., \pi_k$ (called clusters) so that the instances in a cluster are more similar to each other than to instances in other clusters. The number of clusters k also has to be determined.
3. Partition the set $\mathbf{R}$ into **P** variants of disjoint sub-collections of clusters $\begin{matrix} V_1 : \pi_{11}, \pi_{12}, ..., \pi_{1k} \\ V_2 : \pi_{21}, \pi_{22}, ..., \pi_{2l} \\ V_p : \pi_{p1}, \pi_{p1}, ..., \pi_{pr} \end{matrix}$ and to ranging these variants as preset for measure.

Natural steps to approach the two above-mentioned problems are:

Step 1. Embed the instances and the query into a metric space.

Step 2. Handle problems 1 and 2 above as problems concerning points in the metric space.

For instance, a vector space model may maps instances into vectors in a finite dimensional Euclidean space, i.e., let the vector space is of dimension $n =$ 17, and we will be building vectors in $\mathbf{R}_{17}$.

One can expect sparse high dimensional vectors (this is indeed the case in many applications) [Kogan, 2007].

There are several studies which offer better solutions to problems 1 and 2 [Kogan, 2007].

In this work our attention is focused on the problem 3.

As a rule, clustering algorithms have applied their own classification schemes to find a proximity between the instances and to create relevant clustering structures. In other hand, the specific algorithms can lead to different results depending on the baseline to its parameters. That makes it necessary to compare the results in accordance with specified criteria and to rank clustering variants [Stein et al., 2003].

The complexity of analysis increases with the number of instances and clusters, and especially with higher dimensions of the data. Usually in such cases the representatives of the clusters are selected (one or more instances) by which comparing may be done [Stein and Oliver, 1999]. This determines the importance of choosing representatives [Kogan, 2007]. Example of selection representatives is using of centroids of the clusters [Dutot et al. 2011].

The approach we use in this work is based on the methods of abstraction and the construction of hierarchical data structures. These are the Multi-layer Growing Pyramidal Networks (MPGN) realized in system INFOS and presented in [Mitov, 2011]

"***Pyramid***" is a structure that can be described by an oriented graph (Fig. 1), in which:

- there is at least one knot, which does not have outgoing arcs (called vertex of the pyramid);
- all other knots of the graph have at least one outgoing arcs;
- knots which have no incoming arcs (i.e. have no children) are called terminal knots of the pyramid.

Significant difference between pyramidal and tree structures is that the pyramidal structure may have more than one vertex and there is possibility to have more than one way from vertex to terminal node. It is possible to have no way from any vertex to any terminal node, but each terminal node has to reach at least one vertex. Because of this usually the pyramidal structures are called "pyramidal networks" or "growing pyramidal networks" (GPN).

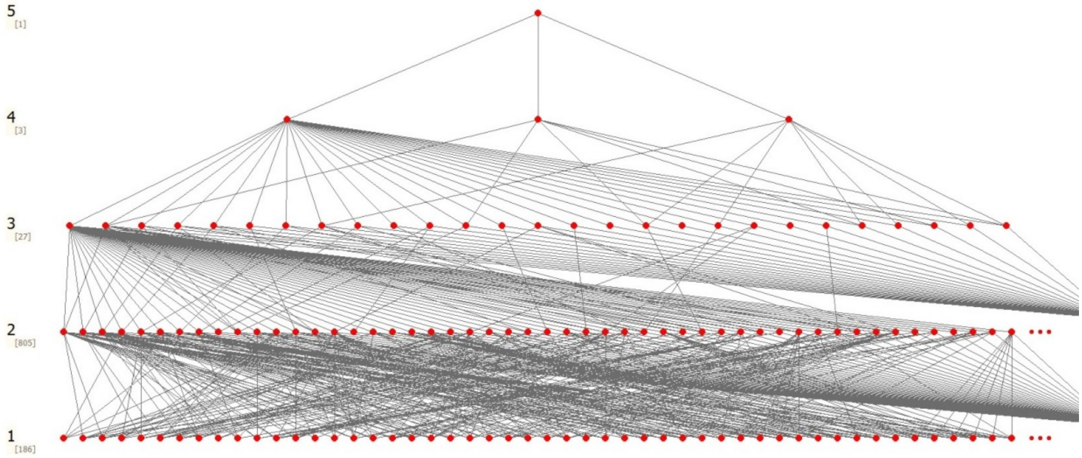


Figure 1. Example of five layer pyramidal structure

Input Data

One possible approach to handle the sparse high dimensional vectors is the Multi-layer Growing Pyramidal Networks (MPGN) realized in system INFOS and presented in [Mitov, 2011]. In this work we use this approach for multi-variant clustering high dimensional data. We will illustrate this by an implementation of MPGN for discovering regularities in data received by National Scientific Center “Institute of mechanization and electrification of agriculture” of Ukrainian Academy of Agriculture Sciences. The observations had collected high dimensional data about wheat crop, including data about fertilizing, weather, water reserves in the top layer of earth, temperature, wind, etc.

In our example we will use a small part of this data to illustrate the idea. In further work it may be extended to whole number of features. The extracted data set from main data collection contains data from 252 real observations of the fertilizing and the corresponded crop of the wheat provided in black earth regions Ukraine, which are rich of humus. Three kinds of fertilizers were chosen: nitric (N), phosphorus (P) and potassium (K) and four selected varieties of wheat – Caucasus, Mironov Jubilee, Mironov 808 and Kharkov 81 (Table 1).

Table 1. Observations of the fertilizing and the corresponded crop of the wheat

variants				Caucasus	Mironov Jubilee					Mironov 808		Kharkov 81						
n	N	P	K	1972	1971	1974	1975	1976	1977	1973	1978	1979	1980	1981	1982	1983	1984	1985
1	0	0	0	24.6	35.0	31.5	24.9	48.0	27.8	24.6	28.8	23.3	33.4	25.1	15.2	21.6	7.1	25.3
2	0.6	0.6	0	29.2	40.6	42.1	24.5	58.8	34.8	42.8	42.7	31.9	33.7	40.2	29.4	39.9	10.0	32.6
3	1.2	1.2	0	-	-	-	-	58.0	36.6	-	50.6	31.2	32.7	47.1	38.5	41.5	11.9	49.0
4	0	0.6	0.3	24.0	40.2	37.0	24.4	46.7	32.3	38.0	26.9	25.2	38.1	30.4	16.2	22.3	7.2	25.6
5	0.6	0.6	0.3	26.5	43.8	32.2	29.5	57.7	32.9	42.6	42.2	32.4	35.5	42.3	29.9	36.7	9.9	31.0
6	0.9	0.6	0.3	26.5	44.2	45.7	31.4	61.3	33.1	41.6	50.6	32.8	35.7	47.4	32.9	39.8	10.2	36.6
7	1.2	0.6	0.3	26.5	40.4	44.2	30.3	57.9	34.9	40.6	50.6	33.1	34.9	46.8	36.4	43.3	12.4	42.6
8	1.5	0.6	0.3	-	-	-	-	53.0	35.4	-	49.5	32.1	32.7	46.5	41.6	43.7	9.6	41.9
9	0.6	1.2	0.3	29.2	46.2	42.8	28.3	58.6	38.0	43.2	44.5	31.8	37.1	39.4	28.5	35.7	10.9	33.2
10	0.6	0.9	0.3	25.8	42.7	41.9	30.3	60.1	35.3	41.7	44.0	30.3	35.9	40.9	28.4	36.0	14.3	34.4
11	0.6	0	0.3	25.8	32.6	34.4	26.5	46.1	32.1	40.6	40.8	29.7	35.5	36.5	20.5	30.7	8.1	26.4
12	0.6	0.6	0.6	28.8	42.7	43.4	32.4	54.4	32.9	43.6	43.3	30.6	38.0	37.7	31.1	37.0	9.8	33.4
13	0.9	0.9	0.6	-	-	-	-	56.0	40.5	-	49.4	34.1	34.7	46.7	36.1	40.1	12.6	38.0
14	0.9	0.6	0.6	-	-	-	-	59.6	35.5	-	47.9	34.3	37.0	45.0	33.2	38.6	13.0	35.0
15	1.2	1.2	0.6	28.8	-	48.1	27.6	56.6	40.2	43.3	48.3	33.1	32.2	50.5	39.6	44.0	13.7	41.2
16	1.2	0	0.6	24.9	-	33.3	25.2	54.3	31.0	38.7	51.3	31.3	35.2	43.2	28.7	39.6	8.2	31.3
17	0	1.2	0.6	28.0	-	38.3	35.3	44.5	32.2	41.3	27.0	25.0	39.7	28.0	16.1	23.2	7.6	26.2
18	0.6	0.6	0.9	-	-	-	-	53.6	33.2	-	43.9	32.6	37.4	42.9	27.5	34.0	10.5	32.2
19	1.2	1.2	0.9	-	-	-	-	60.4	36.8	-	51.4	34.7	34.3	49.8	36.7	42.6	13.0	43.8

Usually, the research is concerned on the data of every variety separately without relationships with others. Here we will try to analyze all varieties in one data set.

Because of great distribution of the values of the crop for different varieties (shown in Figure 2) the normalization of data was provided. The distribution after normalization is shown on Figure 3. After normalization the values of the crop are in the interval [17.58, 49.41] (before it, the interval was [7.10, 61.30]).

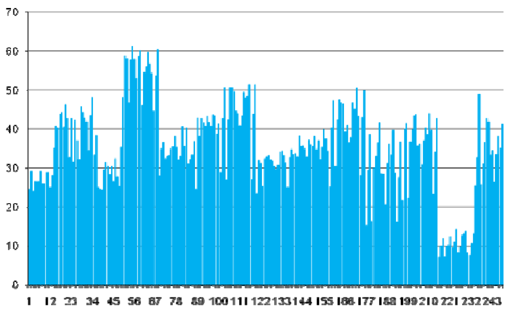


Figure 2. Values of the crop of the different varieties of the wheat before normalization – the vertical interval is [7.10, 61.30]

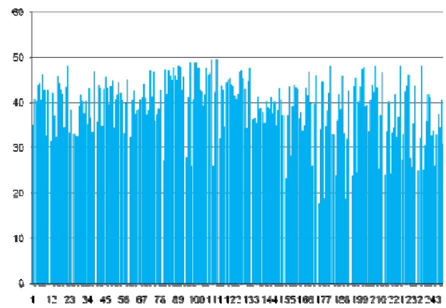


Figure 3. Values of the crop of the different varieties of the wheat after normalization – the vertical interval is [17.58, 49.41]

Variants of clustering

We cluster the data using different kinds of distances between the values of the normalized crop. We provide four different types of clustering:

Case A. One cluster – no distances are used. All instances are assumed to be in this cluster;

Case B. Four clusters based on discretization based on human given intervals. The boundaries are respectively: 35, 40 and 45;

Case C. Five clusters based on discretization realized in system PaGaNe [Mitov et al, 2009a] and especially – the Chi-merge discretization of the normalized crop values [Mitov et al, 2009b];

Case D. Two clusters based on merged clusters from case C: (1+2+3) and (4+5)

The corresponded boundaries of the intervals are presented in Table 2.

Table 2. Boundaries of the intervals for different cases of discretization of the values of the normalized crop of the wheat

Class	Crop normalized	
	min	max
A. One cluster		
1	17.58	49.41
B. Four clusters based on discretization based on human given intervals		
1	17.58	34.99
2	35.00	39.99
3	40.00	44.99
4	45.00	49.41
C. Five clusters based on Chi-merge discretization of the crop values		
1	17.58	23.88
2	24.22	28.07
3	30.43	36.66
4	37.00	43.00
5	43.09	49.41
D. Two clusters based on merging clusters from case C.: (1+2+3) and (4+5)		
1	17.58	36.66
2	37.00	49.41

Case A – one cluster

In the Case A – one cluster (Table 3), at the top of pyramids we receive practically all values used in the experiments. No conclusion may be made directly but, in the same time, this is the starting point for the proposed below pyramidal clustering algorithm.

Table 3. Case A. One cluster – no distances are used. All instances are assumed to be in this cluster

N	P	K	variety
			Caucasus
			Kharkov 81
			Mironov 808
			Mironov jubilee
0N			
0.6N			
0.9N			
1.2N			
	0P		
	0.6P		
	0.9P		
	1.2P		

		0K	
		0.3K	
		0.6K	
		0.9K	

Case B. Four clusters

The Case B corresponds to the human common sense for clustering the data (5 points per interval). The intervals are chosen on the base of understanding that the interesting data are in the top intervals, which were chosen to be equal. The low intervals were merged in one big interval. This way, four intervals were created: (17.58, 34.99), (35.00, 39.99), (40.00, 44.99) and (45.00, 49.41).

This case is more informative (see Table 4). The main conclusion from this case is that the variety "Mironov 808" gives most good crop if the fertilizing is in any of the combinations in class 4. "Mironov jubilee" and "Caucasus" as a rule have middle values of crop. The worst values belong to "Kharkov 81".

Table 4. Case B. Four clusters based on discretization based on human given intervals. The boundaries are respectively: 35, 40 and 45

Class	N	P	K	variety
1	0.6N	0.9P	0.3K	Kharkov 81
1	0.6N	0.6P	0.9K	Kharkov 81
1	0.9N	0.6P	0.3K	Kharkov 81
1	0.9N	0.6P	0.6K	Kharkov 81
1	1.2N	0P	0.6K	Kharkov 81
1	1.2N	1.2P	0.6K	Kharkov 81
1	0N	0P	0K	Mironov 808
1	0N	0.6P	0.3K	Mironov 808
1	1.2N	0P	0.6K	Mironov jubilee

Class	N	P	K	variety
3	0N	1.2P	0.6K	Caucasus
3	0.6N	0.6P	0K	Caucasus
3	0.6N	0.6P	0.6K	Caucasus
3	0.6N	1.2P	0.3K	Caucasus
3	1.2N	1.2P	0.6K	Caucasus
3	0N	0.6P	0.3K	Mironov 808
3	0.6N	0.6P	0.9K	Mironov 808
3	0.6N	0.9P	0.3K	Mironov jubilee
3	0.9N	0.6P	0.6K	Mironov jubilee
3	1.2N	0.6P	0.3K	Mironov jubilee
3	1.2N	1.2P	0.9K	Mironov jubilee
3	1.2N	1.2P	0K	Mironov jubilee

Class	N	P	K	variety
2	0N	0P	0K	Caucasus
2	0N	0.6P	0.3K	Caucasus
2	0.6N	0P	0.3K	Caucasus
2	0.6N	0.6P	0.3K	Caucasus
2	0.6N	0.9P	0.3K	Caucasus
2	0.9N	0.6P	0.3K	Caucasus
2	1.2N	0.6P	0.3K	Caucasus
2	1.2N	0P	0.6K	Caucasus
2	0.6N	0.6P	0.9K	Mironov jubilee

Class	N	P	K	variety
4	0.9N	0.6P	0.3K	Mironov 808
4	0.9N	0.6P	0.6K	Mironov 808
4	0.9N	0.9P	0.6K	Mironov 808
4	1.2N	1.2P	0.6K	Mironov 808
4	1.2N	1.2P	0K	Mironov 808
4	1.2N	1.2P	0.9K	Mironov 808
4	1.5N	0.6P	0.3K	Mironov 808

After the pruning, no generalized patterns exist and, maybe, some important regularity is not discovered. Because of this we continue the experiment with two other cases.

Case C. Five clusters based on the Chi-merge discretization

The Case C is based on discretizator realized in the system PaGaNe [Mitov et al, 2009a] and especially – the Chi-merge discretization of the normalized crop values. In general, pyramidal classifier trained on data preprocessed by Chi-merge achieves lower classification error than those trained on data preprocessed by the other discretization methods. The main reason for this is that using Chi-square statistical measure as criterion for class dependency in adjacent intervals of a feature leads to forming good separating which is convenient for the pyramidal algorithms [Mitov et al, 2009b].

The crop values presented in Table 1 were discretized in five intervals based on the Chi-square statistical measure, respectively (17.58, 23.88), (24.22, 28.07), (30.43, 36.66), (37.00, 43.00), and (43.09, 49.41).

In Table 5 the results of clustering in the Case C are presented.

Table 5. Results from Case C of clustering

Class	N	P	K	variety	Crop
1	0	0	0	Kharkov81 1982	17.58
1	0	1.2	0.6	Kharkov81 1982	18.62
1	0	0.6	0.3	Kharkov81 1982	18.73
1	0	0	0	Kharkov81 1981	23.18
1	0	0	0	Kharkov81 1983	23.61
1	0.6	0	0.3	Kharkov81 1982	23.70
1	0	0	0	Kharkov81 1984	23.88

Class	N	P	K	variety	Crop
2	0	0.6	0.3	Kharkov81 1984	24.22
2	0	0.6	0.3	Kharkov81 1983	24.38
2	0	0	0	Kharkov81 1985	24.84
2	0	0.6	0.3	Kharkov81 1985	25.13
2	0	1.2	0.6	Kharkov81 1983	25.36
2	0	1.2	0.6	Kharkov81 1984	25.56
2	0	1.2	0.6	Kharkov81 1985	25.72
2	0	1.2	0.6	Kharkov81 1981	25.85
2	0	0.6	0.3	Mironov 808 1978	25.86
2	0.6	0	0.3	Kharkov81 1985	25.92
2	0	1.2	0.6	Mironov 808 1978	25.95
2	0	0	0	Mironov 808 1973	27.14
2	0.6	0	0.3	Kharkov81 1984	27.25
2	1.2	0	0.6	Kharkov81 1984	27.58
2	0	0	0	Mironov 808 1978	27.69
2	0	0.6	0.3	Kharkov81 1981	28.07

Class	N	P	K	variety	Crop
3	0.6	0.6	0.3	Kharkov81 1985	30.43
3	1.2	0	0.6	Kharkov81 1985	30.73
3	0	0	0	Mironov jubilee 1974	31.50
3	0.6	0.6	0.9	Kharkov81 1985	31.61
3	0.6	0.6	0.9	Kharkov81 1982	31.80
3	0	0	0	Kharkov81 1979	31.89
3	0.6	0.6	0	Kharkov81 1985	32.00
3	0.6	0.6	0.3	Mironov jubilee 1974	32.20
3	0	0	0	Mironov jubilee 1977	32.24
3	1.5	0.6	0.3	Kharkov81 1984	32.29
3	0	0.6	0.3	Mironov jubilee 1975	32.35

4	0.6	0.6	0	Mironov jubilee 1977	40.36
4	1.2	0.6	0.3	Mironov jubilee 1971	40.40
4	1.2	0	0.6	Mironov jubilee 1976	40.41
4	1.2	1.2	0.6	Kharkov 81 1985	40.44
4	0.6	0.6	0.6	Kharkov 81 1983	40.45
4	1.2	0.6	0.3	Mironov jubilee 1977	40.47
4	0.6	0.6	0.9	Kharkov 81 1980	40.49
4	0.6	0.6	0.6	Mironov jubilee 1976	40.49
4	0.6	0.6	0.3	Mironov 808 1978	40.57
4	0.6	0.6	0	Mironov jubilee 1971	40.60
4	0.6	0	0.3	Kharkov 81 1979	40.65
4	0.6	0.9	0.3	Mironov jubilee 1977	40.94
4	0.6	0.6	0	Mironov 808 1978	41.05
4	1.5	0.6	0.3	Mironov jubilee 1977	41.05
4	1.5	0.6	0.3	Kharkov 81 1985	41.13
4	0.6	0.6	0.6	Kharkov 81 1980	41.13
4	0	1.2	0.6	Caucasus 1972	41.17
4	0.9	0.6	0.6	Mironov jubilee 1977	41.17
4	0	0.6	0.3	Kharkov 81 1980	41.24
4	0.6	0.9	0.3	Kharkov 81 1979	41.48
4	0.9	0.6	0.6	Kharkov 81 1981	41.55
4	0.6	0.6	0.6	Mironov 808 1978	41.62
4	0.9	0.6	0.3	Mironov jubilee 1975	41.63
4	0.9	0.9	0.6	Mironov jubilee 1976	41.68
4	1.2	0.6	0.3	Kharkov 81 1984	41.71
4	0.9	0.9	0.6	Kharkov 81 1982	41.74
4	1.2	0.6	0.3	Kharkov 81 1985	41.82
4	0.6	0.6	0.6	Kharkov 81 1979	41.89
4	0.6	0.9	0.3	Mironov jubilee 1974	41.90
4	0	0.6	0.3	Mironov 808 1973	41.92
4	1.2	0.6	0.3	Kharkov 81 1982	42.09
4	0.6	0.6	0	Mironov jubilee 1974	42.10
4	1.2	1.2	0.6	Mironov jubilee 1976	42.13
4	0.9	0.6	0.6	Kharkov 81 1983	42.20
4	0.6	0.6	0.9	Mironov 808 1978	42.20
4	0.6	0.9	0.3	Mironov 808 1978	42.30
4	0.6	0.6	0.6	Caucasus 1972	42.34
4	1.2	1.2	0.6	Caucasus 1972	42.34
4	0.9	0.9	0.6	Kharkov 81 1984	42.38

3	0.6	0.6	0	Mironov jubilee 1975	32.48
3	0.6	1.2	0.3	Kharkov81 1985	32.59
3	0.6	0	0.3	Mironov jubilee 1971	32.60
3	0.6	0.6	0.6	Kharkov81 1985	32.79
3	0.6	0.9	0.3	Kharkov81 1982	32.84
3	0.6	1.2	0.3	Kharkov81 1982	32.95
3	0.6	0.6	0.6	Kharkov81 1984	32.96
3	0	0	0	Mironov jubilee 1975	33.01
3	0	1.2	0.6	Mironov jubilee 1976	33.12
3	1.2	0	0.6	Kharkov81 1982	33.18
3	1.2	0	0.6	Mironov jubilee 1974	33.30
3	0.6	0.6	0.3	Kharkov81 1984	33.30
3	1.2	0	0.6	Mironov jubilee 1975	33.41
3	0.6	0	0.3	Kharkov81 1983	33.56
3	0.6	0.6	0	Kharkov81 1984	33.64
3	0.6	0	0.3	Kharkov81 1981	33.70
3	0.6	0.9	0.3	Kharkov81 1985	33.77
3	0.6	0.6	0	Kharkov81 1982	33.99
3	0	1.2	0.6	Kharkov81 1979	34.22
3	0.9	0.6	0.3	Kharkov81 1984	34.31
3	0.6	0	0.3	Mironov jubilee 1976	34.31
3	0.9	0.6	0.6	Kharkov81 1985	34.36
3	0.6	0	0.3	Mironov jubilee 1974	34.40
3	0	0.6	0.3	Kharkov81 1979	34.49
3	0.6	0.6	0.3	Kharkov81 1982	34.57
3	0	0.6	0.3	Mironov jubilee 1976	34.76
3	0.6	0.6	0.6	Kharkov81 1981	34.81
3	1.2	1.2	0.6	Kharkov81 1980	34.86
3	0	0	0	Mironov jubilee 1971	35.00
3	0.6	0	0.3	Mironov jubilee 1975	35.13
3	0	0.6	0.3	Caucasus 1972	35.29
3	0.6	0.6	0.9	Kharkov81 1984	35.32
3	1.2	1.2	0	Kharkov81 1980	35.40
3	1.5	0.6	0.3	Kharkov81 1980	35.40
3	0	0	0	Mironov jubilee 1976	35.73
3	0.9	0.6	0.3	Kharkov81 1985	35.93
3	1.2	0	0.6	Mironov jubilee 1977	35.95
3	0.6	0.6	0.6	Kharkov81 1982	35.96
3	0	0	0	Kharkov81 1980	36.16
3	0	0	0	Caucasus 1972	36.17
3	0.6	1.2	0.3	Kharkov 81 1981	36.38
3	0.6	0.6	0	Kharkov 81 1980	36.48
3	1.2	1.2	0.6	Mironov jubilee 1975	36.59
3	1.2	0	0.6	Caucasus 1972	36.61
3	0.6	1.2	0.3	Kharkov 81 1984	36.66

Class	N	P	K	variety	Crop
4	0	0.6	0.3	Mironov jubilee 1974	37.00
4	0.6	0.6	0	Kharkov 81 1981	37.12
4	1.2	1.2	0.9	Kharkov 81 1980	37.13
4	0.6	0.6	0.9	Kharkov 81 1983	37.17
4	0.6	0	0.3	Mironov jubilee 1977	37.23
4	0.9	0.9	0.6	Kharkov 81 1985	37.30
4	0	1.2	0.6	Mironov jubilee 1977	37.34
4	0	0.6	0.3	Mironov jubilee 1977	37.46
4	0.6	1.2	0.3	Mironov jubilee 1975	37.52
4	1.2	1.2	0.9	Kharkov 81 1982	42.43
4	1.2	1.2	0	Mironov jubilee 1977	42.45
4	1.2	1.2	0.9	Mironov jubilee 1977	42.68
4	1.2	0	0.6	Mironov 808 1973	42.69
4	0.6	0.9	0.3	Mironov jubilee 1971	42.70
4	0.6	0.6	0.6	Mironov jubilee 1971	42.70
4	1.2	1.2	0	Kharkov 81 1979	42.71
4	0.6	1.2	0.3	Mironov 808 1978	42.78
4	0.6	1.2	0.3	Mironov jubilee 1974	42.80
4	1.2	0	0.6	Kharkov 81 1979	42.84
4	0.6	0.6	0	Caucasus 1972	42.93
4	0.6	1.2	0.3	Caucasus 1972	42.93
4	1.5	0.6	0.3	Kharkov 81 1981	42.94
4	0.6	0.6	0.3	Mironov jubilee 1976	42.95
4	0.6	0.6	0.6	Mironov jubilee 1975	42.96
4	0	1.2	0.6	Kharkov 81 1980	42.97
4	1.2	1.2	0.9	Kharkov 81 1985	43.00

Class	N	P	K	variety	Crop
5	1.2	0.6	0.3	Mironov jubilee 1976	43.09
5	0.9	0.9	0.6	Kharkov 81 1981	43.12
5	1.2	1.2	0	Mironov jubilee 1976	43.17
5	1.2	0.6	0.3	Kharkov 81 1981	43.21
5	1.2	0	0.6	Kharkov 81 1983	43.29
5	0.6	0.6	0.6	Mironov jubilee 1974	43.40
5	1.2	1.2	0	Kharkov 81 1981	43.49
5	0.9	0.6	0.3	Kharkov 81 1983	43.51
5	0.6	1.2	0.3	Kharkov 81 1979	43.53
5	0.6	1.2	0.3	Mironov jubilee 1976	43.62
5	0.6	0.6	0	Kharkov 81 1983	43.62
5	0.6	0.6	0	Kharkov 81 1979	43.67
5	0.9	0.6	0.6	Kharkov 81 1984	43.73
5	1.2	1.2	0.9	Kharkov 81 1984	43.73
5	0.6	0.6	0	Mironov jubilee 1976	43.76
5	0.9	0.6	0.3	Kharkov 81 1981	43.77
5	0.6	0.6	0.3	Mironov jubilee 1971	43.80
5	0.9	0.9	0.6	Kharkov 81 1983	43.84
5	1.5	0.6	0.3	Kharkov 81 1979	43.94
5	0.6	1.2	0.3	Mironov jubilee 1977	44.07
5	0.9	0.6	0.3	Mironov jubilee 1971	44.20
5	1.2	0.6	0.3	Mironov jubilee 1974	44.20
5	0.6	0.6	0.3	Kharkov 81 1979	44.35
5	0.9	0.6	0.6	Mironov jubilee 1976	44.36
5	1.2	1.2	0	Kharkov 81 1982	44.52
5	0.6	0.6	0.9	Kharkov 81 1979	44.62
5	0.6	0.9	0.3	Mironov jubilee 1976	44.73
5	1.2	0.6	0.3	Mironov 808 1973	44.79
5	0.6	0	0.3	Mironov 808 1973	44.79
5	0.9	0.6	0.3	Kharkov 81 1979	44.90
5	1.2	1.2	0.9	Mironov jubilee 1976	44.96
5	1.2	0.6	0.3	Kharkov 81 1979	45.31
5	1.2	1.2	0.6	Kharkov 81 1979	45.31
5	1.2	1.2	0	Kharkov 81 1983	45.37
5	0	1.2	0.6	Mironov 808 1973	45.56
5	0.9	0.6	0.3	Mironov jubilee 1976	45.62
5	0.9	0.6	0.3	Mironov jubilee 1974	45.70

4	0.9	0.9	0.6	Kharkov 81 1980	37.56
4	0.6	0.9	0.3	Kharkov 81 1981	37.77
4	1.2	0.6	0.3	Kharkov 81 1980	37.78
4	0.6	0.9	0.3	Caucasus 1972	37.93
4	0.6	0	0.3	Caucasus 1972	37.93
4	0.9	0.6	0.3	Kharkov 81 1982	38.04
4	1.2	0	0.6	Kharkov 81 1980	38.10
4	0.6	0.6	0.3	Mironov jubilee 1977	38.16
4	0.6	0.6	0.6	Mironov jubilee 1977	38.16
4	0	1.2	0.6	Mironov jubilee 1974	38.30
4	0.9	0.6	0.3	Mironov jubilee 1977	38.39
4	0.9	0.6	0.6	Kharkov 81 1982	38.39
4	0.6	0.6	0.3	Kharkov 81 1980	38.43
4	0.6	0	0.3	Kharkov 81 1980	38.43
4	0.6	0.6	0.9	Mironov jubilee 1977	38.50
4	0.9	0.6	0.3	Kharkov 81 1980	38.65
4	0.6	0.9	0.3	Kharkov 81 1980	38.86
4	0.6	0.6	0.3	Caucasus 1972	38.96
4	0.9	0.6	0.3	Caucasus 1972	38.96
4	1.2	0.6	0.3	Caucasus 1972	38.96
4	0.6	1.2	0.3	Kharkov 81 1983	39.03
4	0.6	0.6	0.3	Kharkov 81 1981	39.06
4	0.6	0.6	0.3	Mironov jubilee 1975	39.11
4	0.6	0	0.3	Mironov 808 1978	39.22
4	0.6	0.9	0.3	Kharkov 81 1983	39.35
4	1.5	0.6	0.3	Mironov jubilee 1976	39.45
4	0.6	0.6	0.9	Kharkov 81 1981	39.61
4	1.2	0	0.6	Kharkov 81 1981	39.89
4	0.6	0.6	0.9	Mironov jubilee 1976	39.89
4	1.2	1.2	0	Kharkov 81 1984	40.03
4	0.9	0.6	0.6	Kharkov 81 1980	40.05
4	0.6	0.6	0.3	Kharkov 81 1983	40.12
4	0.6	1.2	0.3	Kharkov 81 1980	40.16
4	1.2	0.6	0.3	Mironov jubilee 1975	40.17
4	0.6	0.9	0.3	Mironov jubilee 1975	40.17
4	0	0.6	0.3	Mironov jubilee 1971	40.20

5	1.2	1.2	0.6	Kharkov 81 1982	45.79
5	0.9	0.6	0.3	Mironov 808 1973	45.89
5	1.2	1.2	0.9	Kharkov 81 1981	45.98
5	0.6	0.9	0.3	Mironov 808 1973	46.00
5	0.9	0.6	0.6	Mironov 808 1978	46.05
5	1.2	1.2	0.6	Kharkov 81 1984	46.08
5	0.6	1.2	0.3	Mironov jubilee 1971	46.20
5	1.2	1.2	0.6	Mironov 808 1978	46.43
5	1.2	1.2	0.9	Kharkov 81 1983	46.57
5	1.2	1.2	0.6	Mironov jubilee 1977	46.62
5	1.2	1.2	0.6	Kharkov 81 1981	46.63
5	0.9	0.9	0.6	Kharkov 81 1979	46.68
5	0	1.2	0.6	Mironov jubilee 1975	46.80
5	0.9	0.6	0.6	Kharkov 81 1979	46.95
5	0.9	0.9	0.6	Mironov jubilee 1977	46.97
5	0.6	0.6	0.3	Mironov 808 1973	47.00
5	0.6	0.6	0	Mironov 808 1973	47.22
5	1.2	0.6	0.3	Kharkov 81 1983	47.33
5	0.9	0.9	0.6	Mironov 808 1978	47.49
5	1.2	1.2	0.9	Kharkov 81 1979	47.50
5	1.5	0.6	0.3	Mironov 808 1978	47.58
5	0.6	1.2	0.3	Mironov 808 1973	47.66
5	1.2	1.2	0.6	Mironov 808 1973	47.77
5	1.5	0.6	0.3	Kharkov 81 1983	47.77
5	1.2	1.2	0.6	Mironov jubilee 1974	48.10
5	0.6	0.6	0.6	Mironov 808 1973	48.10
5	1.5	0.6	0.3	Kharkov 81 1982	48.10
5	1.2	1.2	0.6	Kharkov 81 1983	48.10
5	0.6	0.9	0.3	Kharkov 81 1984	48.10
5	1.2	1.2	0	Kharkov 81 1985	48.10
5	1.2	1.2	0	Mironov 808 1978	48.64
5	0.9	0.6	0.3	Mironov 808 1978	48.64
5	1.2	0.6	0.3	Mironov 808 1978	48.64
5	1.2	0	0.6	Mironov 808 1978	49.31
5	1.2	1.2	0.9	Mironov 808 1978	49.41

The results given it the Table 5 show that the clustering is not enough to discover regularities in the data. The additional processing of clusters is needed. Using clusters as classes in MPGN, we have built corresponded pyramids for every case, and have made pruning for the cases B, C, and D. This way, in the corresponded cases we received a number of generalized patterns, which are not contradictory between classes.

Such discretization seems to be more informative but the received results are similar to Case B (see Table 6). In the same time, the instances of the class 1 are contradictory to instances of class 2; and two instances from class 2 are contradictory to instances of class 4 and class 5. Because of this we have to remove them from the resulting Table 6; i.e. to make pruning of the instances by removing the contradictory ones. In Table 6, the contradictory instances are given in italic. After the final pruning there are no instances in class 1 (Table 7).

Table 6. Case C. Five clusters based on the Chi-merge discretization before the final pruning

Class	N	P	K	variety
1	*0N*	*0P*	*0K*	*Kharkov 81*
1	*0N*	*0.6P*	*0.3K*	*Kharkov 81*

Class	N	P	K	variety
4	0N	1.2P	0.6K	Caucasus
4	0.6N	0P	0.3K	Caucasus

1	*0N*	*1.2P*	*0.6K*	*Kharkov 81*
1	*0.6N*	*0P*	*0.3K*	*Kharkov 81*

Class	N	P	K	variety
2	*0N*	*0P*	*0K*	*Kharkov 81*
2	*0N*	*0.6P*	*0.3K*	*Kharkov 81*
2	*0N*	*1.2P*	*0.6K*	*Kharkov 81*
2	*0.6N*	*0P*	*0.3K*	*Kharkov 81*
2	1.2N	0P	0.6K	Kharkov 81
2	0N	0P	0K	Mironov 808
2	*0N*	*0.6P*	*0.3K*	*Mironov 808*
2	*0N*	*1.2P*	*0.6K*	*Mironov 808*

Class	N	P	K	variety
3	0N	0P	0K	Kharkov 81
3	0.6N	0.6P	0.9K	Kharkov 81
3	0.6N	0.9P	0.3K	Kharkov 81
3	0.9N	0.6P	0.3K	Kharkov 81
3	0.9N	0.6P	0.6K	Kharkov 81
3	0N	0P	0K	Caucasus
3	0N	0.6P	0.3K	Caucasus
3	1.2N	0P	0.6K	Caucasus
3	0N	0P	0K	Mironov jubilee

4	0.6N	0.6P	0K	Caucasus
4	0.6N	0.6P	0.3K	Caucasus
4	0.6N	0.6P	0.6K	Caucasus
4	0.6N	0.9P	0.3K	Caucasus
4	0.6N	1.2P	0.3K	Caucasus
4	0.9N	0.6P	0.3K	Caucasus
4	1.2N	0.6P	0.3K	Caucasus
4	1.2N	1.2P	0.6K	Caucasus
4	*0N*	*0.6P*	*0.3K*	*Mironov 808*
4	0.6N	0.6P	0.9K	Mironov 808
4	0.6N	0.6P	0.9K	Mironov jubilee
4	1.5N	0.6P	0.3K	Mironov jubilee

Class	N	P	K	variety
5	*0N*	*1.2P*	*0.6K*	*Mironov 808*
5	0.9N	0.6P	0.3K	Mironov 808
5	0.9N	0.6P	0.6K	Mironov 808
5	0.9N	0.9P	0.6K	Mironov 808
5	1.2N	0.6P	0.3K	Mironov 808
5	1.2N	1.2P	0K	Mironov 808
5	1.2N	1.2P	0.6K	Mironov 808
5	1.2N	1.2P	0.9K	Mironov 808
5	1.5N	0.6P	0.3K	Mironov 808

Table 7. Case C. Five clusters based on the Chi-merge discretization after the final pruning

Class	N	P	K	variety
1	-	-	-	-

Class	N	P	K	variety
2	1.2N	0P	0.6K	Kharkov 81
2	0N	0P	0K	Mironov 808

Class	N	P	K	variety
3	0.6N	0.6P	0.9K	Kharkov 81
3	0.6N	0.9P	0.3K	Kharkov 81
3	0.9N	0.6P	0.3K	Kharkov 81
3	0.9N	0.6P	0.6K	Kharkov 81
3	0N	0P	0K	Caucasus
3	0N	0.6P	0.3K	Caucasus
3	1.2N	0P	0.6K	Caucasus
3	0N	0P	0K	Mironov jubilee

Class	N	P	K	variety
4	0N	1.2P	0.6K	Caucasus
4	0.6N	0P	0.3K	Caucasus
4	0.6N	0.6P	0K	Caucasus
4	0.6N	0.6P	0.3K	Caucasus
4	0.6N	0.6P	0.6K	Caucasus
4	0.6N	0.9P	0.3K	Caucasus
4	0.6N	1.2P	0.3K	Caucasus
4	0.9N	0.6P	0.3K	Caucasus

Class	N	P	K	variety
5	0.9N	0.6P	0.3K	Mironov 808
5	0.9N	0.6P	0.6K	Mironov 808
5	0.9N	0.9P	0.6K	Mironov 808
5	1.2N	0.6P	0.3K	Mironov 808
5	1.2N	1.2P	0K	Mironov 808
5	1.2N	1.2P	0.6K	Mironov 808
5	1.2N	1.2P	0.9K	Mironov 808
5	1.5N	0.6P	0.3K	Mironov 808

4	1.2N	0.6P	0.3K	Caucasus
4	1.2N	1.2P	0.6K	Caucasus
4	0.6N	0.6P	0.9K	Mironov 808
4	0.6N	0.6P	0.9K	Mironov jubilee
4	1.5N	0.6P	0.3K	Mironov jubilee

Case D. Two clusters based on merged clusters from case C.

For the Case D we created two clusters based on merging clusters from case C: classes (1+2+3) and classes (4+5) from Table 6. Again, this is based on “the human common sense”. The idea is that the last two classes (4+5) may contain the most of interesting for us regularities. Table 8 presents the result, which was received after the five steps of processing:

- normalization of data of the crop;
- discretization by the PaGaNe discretizer (Chi-merge)
- merging received intervals into two main (1+2+3) and (4+5)
- generalization into two classes separately
- pruning of the contradictory vertexes and instances between classes.

Table 8. Case D. Two clusters based on merged clusters from case C.:

classes (1+2+3) and (4+5) from Table 6

Class 1			
N	**P**	**K**	**variety**
0N	0P	0K	Caucasus
0N	0.6P	0.3K	Caucasus
1.2N	0P	0.6K	Caucasus
0N	0P	0K	Kharkov 81
0.6N	0.6P	0.9K	Kharkov 81
0.6N	0.9P	0.3K	Kharkov 81
0.9N	0.6P	0.3K	Kharkov 81
0.9N	0.6P	0.6K	Kharkov 81
0N	0P	0K	Mironov 808
0N	0.6P	0.3K	Mironov 808
0N	1.2P	0.6K	Mironov 808
0N	0P	0K	Mironov jubilee

Class 2			
N	**P**	**K**	**variety**
0.6N	0.6P	0.3K	Caucasus
0.6N	0.6P	0.6K	Caucasus
0.6N	0.6P	0K	Caucasus
0.6N	0P	0.3K	Caucasus
0.6N	1.2P	0.3K	Caucasus
0N	1.2P	0.6K	Caucasus
1.2N	0.6P	0.3K	Caucasus
1.2N	1.2P	0.6K	Caucasus
1.2N	0.6P	0.3K	Kharkov 81
-	-	-	Mironov 808
0.6N	0.6P	0.6K	Mironov jubilee
0.6N	1.2P	0.3K	Mironov jubilee
1.2N	0.6P	0.3K	Mironov jubilee
1.2N	1.2P	0K	Mironov jubilee
1.5N	0.6P	0.3K	Mironov jubilee

The main conclusion from this case is that the variety “Mironov 808”, “Mironov jubilee”, and “Caucasus” are good with small exception (3 for the first variety, one for the second, and 3 for the third). The worst values belong to “Kharkov 81”.

Let mention the special instance in class 2 for variety Mironov 808 which contains dashes in all positions. This means that all instances of Mironov 808 in class 2 are not contradictory to ones in class 1. Because of this only one generalized instance is given as result. In the same time in class 1 there exist just three instances which have no contradictory to instances of the class 2 and they are shown in the Table 8.

Ranging of the variants

In clustering tasks, the procedure of evaluating the results is known under the term cluster validity [Halkidi et al., 2002]. Every clustering result, regardless the algorithm, can be evaluated whether it has certain structural properties or not.

Method of assessment we use is based on the measure of Dunn [Stein et al., 2003]. It is defined as follows.

Let

- C = {C1, ..., Ck} be a clustering of a set of D,
- δ : C × C → **R** be cluster to cluster distance measure, and
- Δ : *C* → **R** be a cluster diameter measure.

Then all measures ***I*** of the form $I(C) = \frac{\min_{i \neq j} \{\delta(C_i, C_j)\}}{\max_{1 \le l \le k} \{\Delta(C_l)\}}$ are called Dunn indices.

When comparing the different variants of clustering by Dunn indices as the best variant is considered the greatest Dunn index. We use Dunn index to ensure that the pyramidal structures retain the relationship between clusters. The values of Dunn index at the top of the pyramids is given in Table 9 and those in their bases, i.e. the original clusters before building of the pyramids are in Table 10.

Table 9. The values of Dunn index at the top of the pyramids.

Variant A	$I_{VA-dunn} = \frac{0}{6.5901745} = 0$
Variant B	$I_{VB-dunn} = \frac{1.2351518}{9.10439454} = 0.1356655$
Variant C	$I_{VC-dunn} = \frac{1.480675522}{5.14198405} = 0.2879580$
Variant D	$I_{VD-dunn} = \frac{1.48067552}{11.4995826} = 0.1287591$

Table 10. The values of Dunn index at the base of the pyramids.

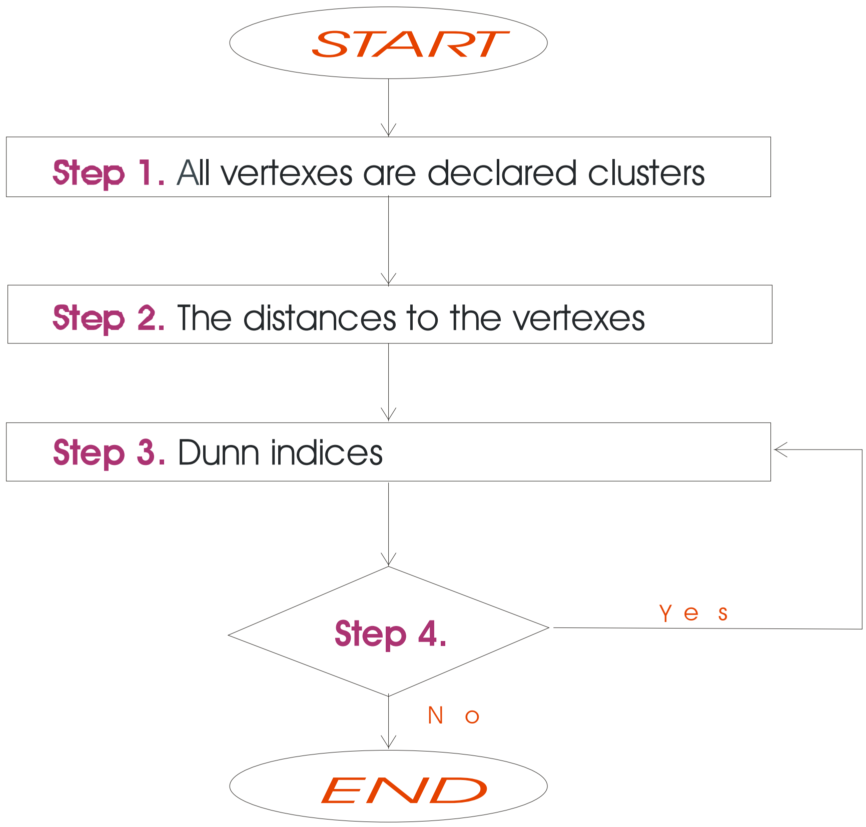
Variant A	$I_{BA-dunn} = \frac{0}{6.5901745} = 0$
Variant B	$I_{BB-dunn} = \frac{0.4534310}{34.9064410} = 0.0129899$
Variant C	$I_{BC-dunn} = \frac{0.3354100}{23.8800000} = 0.0140456$
Variant D	$I_{BD-dunn} = \frac{0.6177380}{36.6857680} = 0,0168386$

The results show that Variant C has higher index at the level of vertexes but Variant D has higher index at the base level. This is due to removing the empty Cluster 1 of Variant C after the pruning.

As a rule the indices at the tops of the pyramids are greater than these at the bases and, in general, the pyramidal structures retain the relationship between clusters.

One algorithm for pyramidal clustering high dimensional data

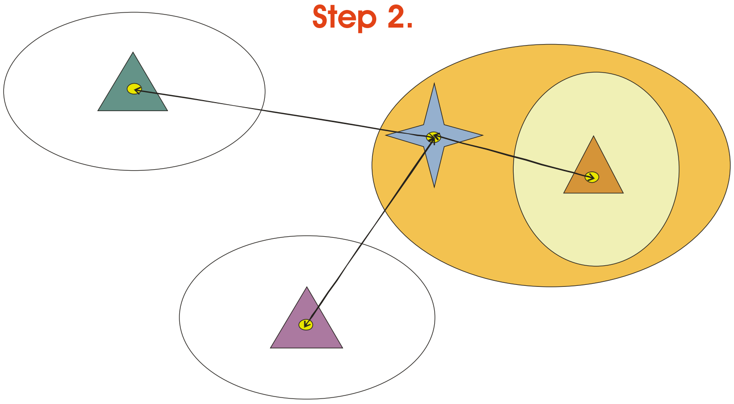
1. Pre-processing to prepare the instances to be suitable for calculating the distances between them (discretization, sorting and numbering, formation fuzzy sets, etc.)
2. Generalization of the instances.
3. As result vertexes with or without predecessors appear.
4. Vertexes without predecessors are single pyramids and do not form clusters. They are declared as regular instances, which should be connected to the corresponding clusters.
5. Some instances can belong to more than one pyramid. Such instances have to be attached only to one pyramid. For this purpose, we use the following algorithm.
6. Algorithm:



Step 1. All vertexes are declared clusters. All instances form a set of candidates for clustering.

Step 2. For each instance from the set of candidates, its distances to the vertexes are calculated. The vertex which is the most close is determined. If the minimum distance is only to one vertex, the instance is attached to its cluster. In other case, the instance remains as element of the set of candidates for further processing.

Step 2.



Step 3. For each instance from the set of candidates: 3a) the diameters of the clusters and 3b) the distances between the clusters are computed using the instance as element of the corresponded clusters.

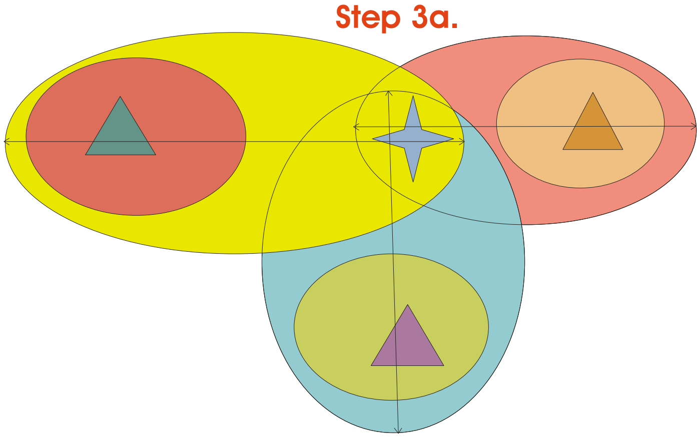
If there exist only one cluster, for which the *Dunn index* is minimally decreased, the instance is attached to this cluster. In other case, the instance remains as element of the set of candidates for further processing.

Definition 3 (Dunn Measure) Let $C = \{C_1, \ldots, C_K\}$ be a clustering of set of objects D, $\delta : C \times C \to R$ be a cluster to cluster distance, and $\Delta : C \to R$ be a cluster diameter measure. Then all measures I of the form

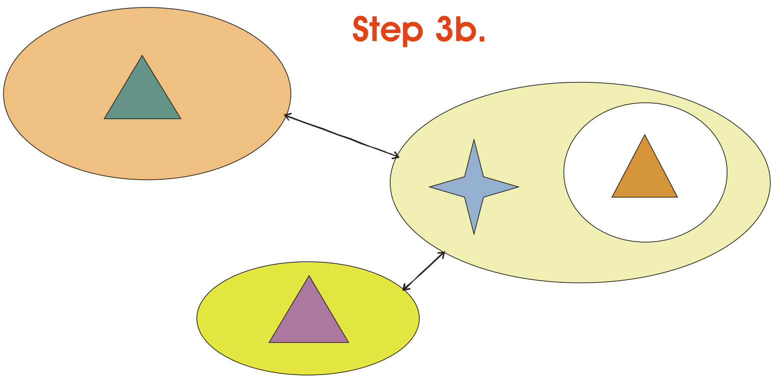
$$I(C) = \frac{\min_{i \neq j}\{\delta(C_i, C_j)\}}{\max_{1 \leq l \leq k}\{\Delta(C_l)\}}$$

are called Dunn indices.

Step 3a.



Step 3b.



Step 4. If the set of candidates for clustering has not been changed after the Step 3, the instances of this set are declared as impossible to be clustered and algorithm stops. Otherwise, it continues from Step 3.

As a result of implementation of the algorithm we obtain the densest clustering.

Conclusion

In this work we have used a small part of data to illustrate a possible clustering approach to handle the sparse high dimensional vectors. The extracted data set from main data collection contained data from 252 real observations of the fertilizing and the corresponded crop of the wheat provided in black earth regions Ukraine, which are rich of humus. Three kinds of fertilizers were chosen: nitric (N), phosphorus (P) and potassium (K) and four varieties of wheat – Caucasus, Mironov Jubilee, Mironov 808 and Kharkov 81.

Our main goal in this work was to illustrate using the approach for multi-variant clustering high dimensional data based on the Multi-layer Growing Pyramidal Networks (MPGN) and the system INFOS. We outlined an implementation of MPGN for discovering regularities in data received by National Scientific Center “Institute of mechanization and electrification of agriculture” of Ukrainian Academy of Agriculture Sciences. The observations had collected high dimensional data about wheat crop, including data about fertilizing, weather, water reserves in the top layer of earth, temperature, wind, etc.

The analysis of the results from different cases permits us to say that the Heady and Dillon advices are still actual (in our example, too). The main theirs advice is not to accept only one equation for characterizing the agricultural production in different conditions [Heady and Dillon, 1961].

Taking in account all cases we may draw inference that the variety “Mironov 808” is stable in all observations. “Mironov jubilee” shows less stability but with proper fertilizing gives good crop. “Caucasus” and “Kharkov 81” could not be recommended to be used. Let remember that our example do not take in account many factors which were observed. In further work, data will be extended to whole number of features. The conclusion may differ when we will use great number of dimensions.

Similar results were received by parallel independent experiments with the same data provided by the program complex CONFOR which is based on pyramidal structures, too.

A possible extension of the investigated area is in direction of fuzzy clustering [Hoeppner et al, 1997]. As it is outlined in [Bodyanskiy et al, 2011] the problem of multidimensional data clusterization is an important part of exploratory data analysis [Tukey, 1977], [Höppner et al, 1999], with its goal of retrieval in the analyzed data sets of observations some groups (classes, clusters) that are homogeneous in some sense. Traditionally, the approach to this problem assumes that each observation may belong to only one cluster, although more natural is the situation where the processed vector of features could refer to several classes with different levels of membership (probability, possibility). This situation is the subject of fuzzy cluster analysis [Bezdek, 1981]; [Gath and Geva, 1989]; [Höppner et al, 1999], which is based on the assumption that the classes of homogeneous data are not separated, but overlap, and each observation can be attributed to a certain level of membership to each cluster, which lies in the range of zero to one [Höppner et al, 1999]. Initial information for this task is a sample of observations, formed from N -dimensional feature x(1),x(2),…,x(k),…,x(N).

The result of clustering is segmentation of the original data set into m classes with some level of membership $u_j(k)$ of k -th feature vector x(k) to j -th cluster, $j=1, \ 2, \cdots, \ m$. [Bodyanskiy et al, 2011]

What we have seen from the experiments is that the multi-variant clustering combined with pyramidal generalization and pruning give reliable results. Using algorithms for fuzzy clustering will give new possibilities.

Bibliography

[Bezdek, 1981] Bezdek J.C. Pattern Recognition with Fuzzy Objective Function Algorithms, N.Y.:Plenum Press., 1981.

[Bodyanskiy et al, 2011] Bodyanskiy Y., Kolchygin B., Pliss I. Adaptive Neuro-fuzzy Kohonen Network with Variable Fuzzifier. International Journal “Information Theories and Applications”, Vol. 18, Number 3, 2011, pp. 215 – 223

[Dutot et al. 2011] Antoine Dutot, Damien Olivier and Guilhelm Savin. CENTROIDS: A DECENTRALIZED APPROACH Proceedings of EPNACS 2011 within ECCS’11 Emergent Properties in Natural and Artificial Complex Systems Vienna, Austria - September 15, 2011,p.23-28

[Gath and Geva, 1989] Gath I., Geva A.B. Unsupervised optimal fuzzy clustering In: Pattern Analysis and Machine Intelligence.,1989., 2., 7., P. 773-787

[Gladun and Vashchenko, 2000] Gladun V.P., Vaschenko N.D. Analytical Processes in Pyramidal Networks. Int. Journal Information Theories and Applications, Vol.7, No.3, 2000, pp.103-109.

[Gladun et al, 2008] Gladun V., Velichko V., Ivaskiv Y. Selfstructurized Systems. International Journal Information Theories and Applications. FOI ITHEA, Sofia, Vol.15,N.1, 2008, pp.5-13.

[Gladun, 1987] Gladun V.P. Planning of Solutions. Kiev, Naukova Dumka, 1987, 168 p, (in Russian).

[Gladun, 1994] Gladun V.P. Processes of New Knowledge Formation. Sofia, SD Pedagog 6, 1994, 192 p, (in Russian).

[Gladun, 2000] Gladun V.P. Partnership with Computers.. Man-Computer Task-oriented Systems. Kiev, Port-Royal, 2000, 120 p, (in Russian).

[Gladun, 2003] Gladun V.P. Intelligent Systems Memory Structuring. Int. Journal Information Theories and Applications, Vol.10, No.1, 2003, pp.10-14.

[Halkidi et al., 2002] Halkidi M., Yannis Batistakis, and Michalis Vazirgiannis. Clustering Validity Checking Methods: Part II. ACM SIGMOD Record, 31(3):19–27, 2002. ISSN 0163-5808.

[Heady and Dillon, 1961] Heady E.O., Dillon J.L. Agricultural Production Functions. Ames, Iowa : Iowa State University Press, 1961. 667 p.

[Hoeppner et al, 1997] Hoeppner F., Klawonn F., Kruse R. Fuzzy-Clusteranalysen. –Braunschweig:Vieweg, 1997. – 280S.

[Höppner et al, 1999] Höppner F., Klawonn F., Kruse R., Runkler T. Fuzzy Clustering Analysis: Methods for Classification, Data Analysis and Image Recognition. Chichester: John Willey & Sons., 1999,

[Kiselyova et al, 1998] Kiselyova N.N., Gladun V.P., Vashchenko N.D., LeClair S.R., Jackson 6.G. Prediction of Inorganic Compounds Perspective for Search of New Electrooptical Materials// Perspektivnie Materiali, 1998, N3. pp.28 -32. (in Russian).

[Kogan, 2007] Jacob Kogan. Introduction to Clustering Large and High-Dimensional Data. Cambridge University Press, UK, 2007. 222 p.

[Kogan, 2007] Kogan J. Introduction to Clustering Large and High-Dimensional Data. Cambridge University Press, UK, 2007. 222 p.

[Luo et al, 2009] Ping Luo, Hui Xiong, Guoxing Zhan, Junjie Wu, Zhongzhi Shi. Information-Theoretic Distance Measures for Clustering Validation: Generalization and Normalization. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 21, NO. 9, SEPTEMBER 2009. Published by the IEEE Computer Society. pp. 1249-1262.

[Markov, 2004] Markov, K.: Multi-domain information model. Int. J. Information Theories and Applications, 11/4, 2004, pp.303-308.

[Mitov et al, 2009a] Mitov, I., Ivanova, K., Markov, K., Velychko, V., Vanhoof. K., Stanchev, P.: "PaGaNe" – A classification machine learning system based on the multidimensional numbered information spaces. In World Scientific Proc. Series on Computer Engineering and Information Science, No.2, pp.279-286.

[Mitov et al, 2009b] Mitov, I., Ivanova, K., Markov, K., Velychko, V., Stanchev, P., Vanhoof, K.: Comparison of discretization methods for preprocessing data for pyramidal growing network classification method. In Int. Book Series Information Science & Computing – Book No: 14. New Trends in Intelligent Technologies, 2009, pp. 31-39.

[Mitov, 2011] Iliya Mitov. Class Association Rule Mining Using Multi-Dimensional Numbered Information Spaces. PhD thesis. Promoters: K. Vanhoof, Kr. Markov. Hasselt University, 2011

[Stein and Oliver, 1999] Stein B., Oliver Niggemann On the Nature of Structure and Its Identifcation Widmayer et al. (Eds.): WG'99, LNCS 1665, pp. 122-134, 1999. C Springer-Verlag Berlin Heidelberg 1999

[Stein et al., 2003] Stein B., Sven Meyer zu Eissen, Frank Wißbrock On Cluster Validity and the Information Need of Users, 3rd IASTED Int. Conference on Artificial Intelligence and Applications (AIA 03). Benalmádena, Spain, September 2003. Edited by M. H. Hanza.

[Tukey, 1977] Tukey J. W. Exploratory Data Analysis. Reading, MA: Addison-Wesley Publ. Company, Inc., 1977.

Appendix 1. Diameters of clusters and distances between clusters for different variants

Below the data for diameters of clusters and distances between clusters are given. The data are computed using SPSS. For every variant the Dunn indices are given.

Table 1. Case A – one cluster

A.1.1.				42.93	Caucasus
A.1.2.				48.10	Kharkov 81
A.1.3.				49.41	Mironov 808
A.1.4.				48.10	Mironov jubilee
A.1.5.	0N			46.80	
A.1.6.	0.6N			48.10	
A.1.7.	0.9N			48.64	
A.1.8.	1.2N			49.41	
A.1.9.		0P		49.31	
A.1.10.		0.6P		48.64	
A.1.11.		0.9P		48.10	
A.1.12.		1.2P		49.41	
A.1.13.			0K	48.64	
A.1.14.			0.3K	48.64	
A.1.15.			0.6K	49.31	
A.1.16.			0.9K	49.41	

CASE A.1.

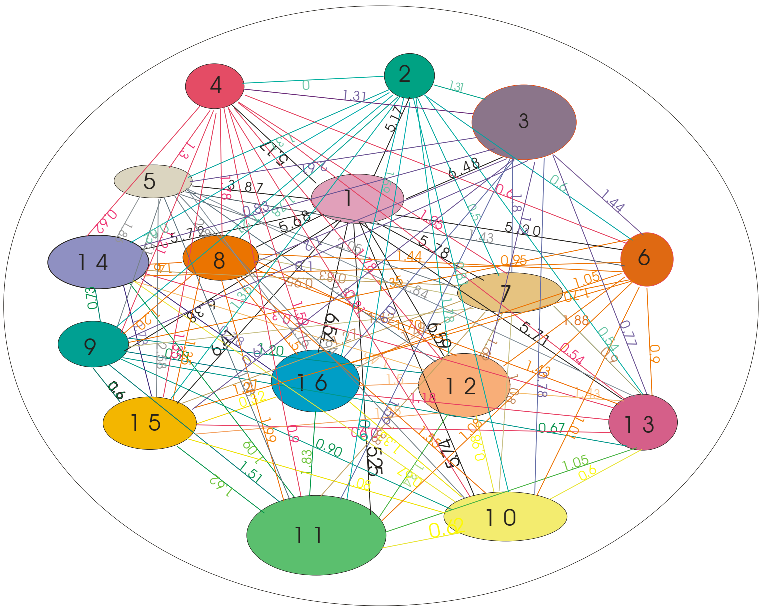


Table 2. Case A. Internal distances and diameter

	A.1.1.	A.1.2.	A.1.3.	A.1.4.	A.1.5.	A.1.6.	A.1.7.	A.1.8.	A.1.9.	A.1.10.	A.1.11.	A.1.12.	A.1.13.	A.1.14.	A.1.15.	A.1.16.
A.1.1.	0	5.17	6.48	5.17	3.87	5.20	5.78	6.5901745	6.38	5.74	5.23	6.5901745	5.71	5.72	6.41	6.54
A.1.2.		0	1.31	0	1.3	0.6	1.05	1.77654158	1.21	0.81	0.9	1.77654158	0.54	0.63	1.35	1.59
A.1.3.			0	1.31	2.61	1.44	1.18	1.2	0.1	0.98	1.59	1.2	0.77	0.83	0.61	0.9
A.1.4.				0	1.3	0.6	1.05	1.77654158	1.21	0.81	0.9	1.77654158	0.54	0.62	1.35	1.59
A.1.5.					0	1.43	2.05	2.87264686	2.51	1.94	1.58	2.87264686	1.84	1.86	2.58	2.76
A.1.6.						0	1.05	1.44086779	1.35	1.01	1.08	1.87512666	0.81	0.86	1.48	1.699
A.1.7.							0	0.82637764	1.12	1.08	1.38	1.68609015	0.9	0.95	1.27	1.49
A.1.8.								0	1.20	1.55	1.99	1.69705627	1.43	1.46	1.35	1.5
A.1.9.									0	0.899	1.51	1.20415946	0.67	0.73	0.6	0.91
A.1.10.										0	0.62	0.97616597	0.6	0.67	1.08	1.33
A.1.11.											0	1.3439122	1.05	1.09	1.62	1.83
A.1.12.												0	1.43	1.46	1.35	1.5
A.1.13.													0	0.3	0.899	1.18
A.1.14.														0	0.73	0.98
A.1.15.															0	0.316
A.1.16.																0

$$I_{dunn} = \frac{0}{6.5901745} = 0$$

Table 3. Case B. Four clusters

	Class	N	P	K	Crop	variety
B.1.1.	1	0.6N	0.9P	0.3K	32.84	Kharkov 81
B.1.2.	1	0.6N	0.6P	0.9K	31.80	Kharkov 81
B.1.3.	1	0.9N	0.6P	0.3K	34.31	Kharkov 81
B.1.4.	1	0.9N	0.6P	0.6K	34.36	Kharkov 81
B.1.5.	1	1.2N	0P	0.6K	27.58	Kharkov 81
B.1.6.	1	1.2N	1.2P	0.6K	34.86	Kharkov 81
B.1.7.	1	0N	0P	0K	27.69	Mironov 808
B.1.8.	1	0N	0.6P	0.3K	25.86	Mironov 808
B.1.9.	1	1.2N	0P	0.6K	33.41	Mironov jubilee

	Class	N	P	K	Crop	variety
B.3.1.	3	0N	1.2P	0.6K	41.17	Caucasus
B.3.2.	3	0.6N	0.6P	0K	42.93	Caucasus
B.3.3.	3	0.6N	0.6P	0.6K	42.34	Caucasus
B.3.4.	3	0.6N	1.2P	0.3K	42.93	Caucasus
B.3.5.	3	1.2N	1.2P	0.6K	42.34	Caucasus
B.3.6.	3	0N	0.6P	0.3K	41.92	Mironov 808
B.3.7.	3	0.6N	0.6P	0.9K	43.9	Mironov 808
B.3.8.	3	0.6N	0.9P	0.3K	44.73	Mironov jubilee
B.3.9.	3	0.9N	0.6P	0.6K	44.36	Mironov jubilee
B.3.10.	3	1.2N	0.6P	0.3K	44.20	Mironov jubilee
B.3.11.	3	1.2N	1.2P	0.9K	44.96	Mironov jubilee
B.3.12.	3	1.2N	1.2P	0K	43.17	Mironov jubilee

	Class	N	P	K	Crop	variety
B.2.1.	2	0N	0P	0K	36.17	Caucasus
B.2.2.	2	0N	0.6P	0.3K	35.29	Caucasus
B.2.3.	2	0.6N	0P	0.3K	37.93	Caucasus
B.2.4.	2	0.6N	0.6P	0.3K	38.96	Caucasus
B.2.5.	2	0.6N	0.9P	0.3K	38.96	Caucasus
B.2.6.	2	0.9N	0.6P	0.3K	38.96	Caucasus
B.2.7.	2	1.2N	0.6P	0.3K	38.96	Caucasus
B.2.8.	2	1.2N	0P	0.6K	36.61	Caucasus
B.2.9.	2	0.6N	0.6P	0.9K	38.50	Mironov jubilee

	Class	N	P	K	Crop	variety
B.4.1.	4	0.9N	0.6P	0.3K	45.89	Mironov 808
B.4.2.	4	0.9N	0.6P	0.6K	46.05	Mironov 808
B.4.3.	4	0.9N	0.9P	0.6K	47.49	Mironov 808
B.4.4.	4	1.2N	1.2P	0.6K	47.77	Mironov 808
B.4.5.	4	1.2N	1.2P	0K	48.64	Mironov 808
B.4.6.	4	1.2N	1.2P	0.9K	49.41	Mironov 808
B.4.7.	4	1.5N	0.6P	0.3K	47.58	Mironov 808

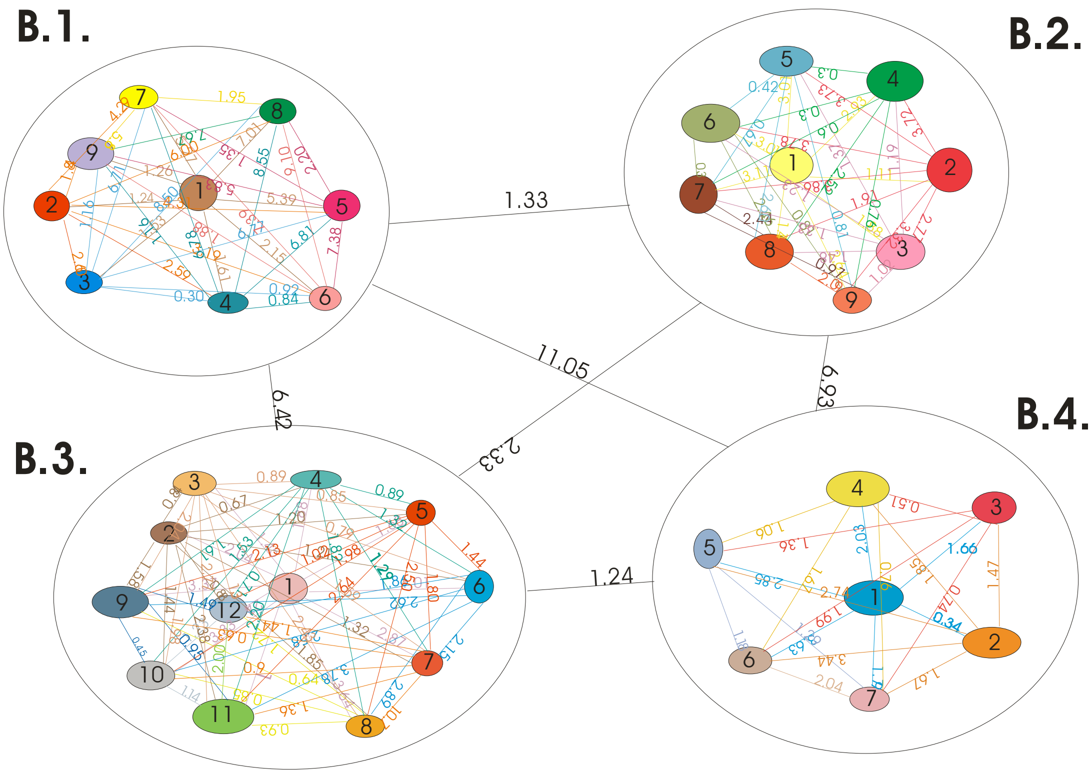


Table 4. Case B. Internal distances and diameters

	B.1.1.	B.1.2.	B.1.3.	B.1.4.	B.1.5.	B.1.6.	B.1.7.	B.1.8.	B.1.9.
B.1.1.	0	1.2375783	1.53	1.6063624	5.37843844	2.14951157	5.27091074	7.01216087	1.25892812
B.1.2.		0	2.59809546	2.5949181	4.3149044	3.18960813	4.29209739	6.00029999	1.84447825
B.1.3.			0	0.3041381	6.77	0.91787799	6.71449179	8.49779383	1.161895
B.1.4.				0	6.81310502	0.83666003	6.78372317	8.55277733	1.16297033
B.1.5.					0	7.37823827	1.34614264	2.20190826	5.83
B.1.6.						0	7.39248943	9.10439454	1.88215302
B.1.7.							0	1.9490767	5.87523617
B.1.8.								0	7.6741449
B.1.9.									0

	B.2.1.	B.2.2.	B.2.3.	B.2.4.	B.2.5.	B.2.6.	B.2.7.	B.2.8.	B.2.9.
B.2.1.	0	1.1065261	1.88350737	2.93156955	3.00734102	3.00734102	3.11032153	1.41194901	2.63797271
B.2.2.		0	2.7730128	3.7187229	3.7308042	3.77874318	3.86120448	1.90588562	3.32025601
B.2.3.			0	1.1920151	1.36780847	1.22918672	1.33450365	1.48067552	1.0222035
B.2.4.				0	0.3	0.3	0.6	2.51644591	0.75604233
B.2.5.					0	0.42426407	0.67082039	2.60432333	0.81338798
B.2.6.						0	0.3	2.46221445	0.81338798
B.2.7.							0	2.44386988	0.96519428
B.2.8.								0	2.0933466
B.2.9.									0

	B.3.1.	B.3.2.	B.3.3.	B.3.4.	B.3.5.	B.3.6.	B.3.7.	B.3.8.	B.3.9.	B.3.10.	B.3.11.	B.3.12.
B.3.1.	0	2.04392	1.4453	1.8835	1.6760	1.0062	2.87452605	3.63505158	3.3683	3.3273	3.98674052	2.4083
B.3.2.		0	0.8415	0.6708	1.1950	1.21248	1.32321578	1.8493242	1.5795	1.4363	2.37716217	0.8818
B.3.3.			0	0.8934	0.8485	0.7916	1.58858428	2.42736483	2.0422	1.9773	2.77027074	1.33
B.3.4.				0	0.89336443	1.3191	1.28875909	1.82482876	1.6078	1.5274	2.20020454	0.7125
B.3.5.					0	1.4375	1.80099972	2.50041996	2.1285	1.9773	2.63711964	1.0242
B.3.6.						0	2.15415877	2.88896175	2.6179	2.5765	3.37662553	1.8581
B.3.7.							0	1.06719258	0.6258	0.9	1.35779233	1.4363
B.3.8.								0	0.6379	0.8549	0.92892411	1.7244
B.3.9.									0	0.4534	0.9486833	1.4920
B.3.10.										0	1.13912247	1.2292
B.3.11.											0	2.0035
B.3.12.												0

	B.4.1.	B.4.2.	B.4.3.	B.4.4.	B.4.5.	B.4.6.	B.4.7.
B.4.1.	0	0.34	1.65529454	2.01851431	2.84648907	3.63323547	1.79334882
B.4.2.		0	1.47091808	1.84618526	2.74191539	3.43941856	1.6705987

B.4.3.			0	0.5083306	1.36473441	1.98907013	0.74033776
B.4.4.				0	1.0568349	1.66721324	0.75901252
B.4.5.					0	1.1844408	1.28980619
B.4.6.						0	2.03933813
B.4.7.							0

Table 5. Case B. External distances between clusters

	B.2.1.	B.2.2.	B.2.3.	B.2.4.	B.2.5.	B.2.6.	B.2.7.	B.2.8.	B.2.9.
B.1.1.	3.5141002	2.5401772	5.1689554	6.1273485	6.12	6.1346883	6.15665494	3.9335607	5.699614
B.1.2.	4.5416847	3.5916709	6.1884489	7.1850957	7.1913559	7.1913559	7.21010402	4.8934752	6.7
B.1.3.	2.172464	1.3305638	3.6816301	4.6596674	4.6693147	4.65	4.65966737	2.4145393	4.2433595
B.1.4.	2.1922819	1.3284954	3.6448457	4.6195238	4.6292548	4.6097722	4.61952378	2.3478714	4.1616824
B.1.5.	8.6941417	7.831609	10.371716	11.415533	11.435226	11.403701	11.3997544	9.03	10.957025
B.1.6.	2.22623	1.4404513	3.3637628	4.1976184	4.1653331	4.1653331	4.15451562	2.1219095	3.7496133
B.1.7.	8.48	7.6295478	10.261949	11.305879	11.325763	11.325763	11.3535413	9.0203326	10.880538
B.1.8.	10.3318	9.43	12.099789	13.113733	13.117164	13.13088	13.154847	10.83755	12.668449
B.1.9.	3.0688108	2.3290341	4.5695076	5.6224994	5.6623758	5.5984373	5.59039355	3.2	5.1689554

	B.3.1.	B.3.2.	B.3.3.	B.3.4.	B.3.5.	B.3.6.	B.3.7.	B.3.8.	B.3.9.	B.3.10.	B.3.11.	B.3.12.
B.1.1.	8.3623501	10.098916	9.509469	10.094459	9.5283787	9.104746	11.0803249	11.89	11.531713	11.379789	12.15337	10.356104
B.1.2.	9.4131238	11.166329	10.544269	11.162298	10.578355	10.155511	12.1	12.94739	12.567164	12.428998	13.187327	11.437084
B.1.3.	6.9512301	8.6304345	8.0412002	8.6460627	8.0635538	7.6630346	9.61343331	10.428634	10.054477	9.894549	10.687961	8.8904218
B.1.4.	6.895368	8.5962143	7.9856371	8.6014476	8.0081459	7.6192913	9.5494293	10.38301	10	9.8491421	10.625441	8.8558512
B.1.5.	13.69555	15.385139	14.78437	15.411441	14.8087	14.405749	16.3447973	17.186695	16.793403	16.633532	17.423961	15.647623
B.1.6.	6.4230912	8.1366394	7.5279745	8.097833	7.48	7.1926073	9.08469042	9.8973178	9.5236548	9.3640589	10.104454	8.3316325
B.1.7.	13.546601	15.263604	14.686814	15.301882	14.760166	14.245803	16.2571246	17.076932	16.715828	16.567139	17.376504	15.572745
B.1.8.	15.324689	17.083176	16.493647	17.091077	16.537243	16.06	18.0599446	18.88192	18.524308	18.379217	19.156461	17.364507
B.1.9.	7.9433998	9.5765547	8.970223	9.6187525	9.0102664	8.6203306	10.5285374	11.375518	10.970529	10.810833	11.616045	9.8517816

	B.4.1.	B.4.2.	B.4.3.	B.4.4.	B.4.5.	B.4.6.	B.4.7.
B.1.1.	13.056895	13.220216	14.656142	14.948073	15.817079	16.594424	14.7704976
B.1.2.	14.10596	14.256314	15.698602	15.99534	16.885366	17.630431	15.8170288
B.1.3.	11.58	11.743832	13.186827	13.480045	14.348829	15.126797	13.2835575
B.1.4.	11.533902	11.69	13.133427	13.426768	14.308333	15.06793	13.2370087
B.1.5.	18.32474	18.482178	19.932589	20.22563	21.102692	21.865015	20.0134954
B.1.6.	11.054452	11.210089	12.637124	12.91	13.793056	14.553092	12.7412087
B.1.7.	18.234583	18.401619	19.849937	20.160516	21.018623	21.804779	19.9577579
B.1.8.	20.050209	20.212276	21.652873	21.953089	22.821446	23.595815	21.771734
B.1.9.	12.501616	12.657788	14.111924	14.410052	15.28898	16.047741	14.1890415

	B.3.1.	B.3.2.	B.3.3.	B.3.4.	B.3.5.	B.3.6.	B.3.7.	B.3.8.	B.3.9.	B.3.10.	B.3.11.	B.3.12.
B.2.1.	5.1768716	6.8130463	6.2569082	6.8983766	6.4272	5.7889982	7.82833954	8.6332844	8.2828799	8.1468337	8.9974496	7.2027772
B.2.2.	5.9181416	7.6693937	7.081843	7.686976	7.1827919	6.63	8.65171081	9.4638047	9.1194792	8.9904449	9.781048	7.9990249
B.2.3.	3.5196023	5.0447993	4.4607286	5.1419841	4.6193181	4.0792279	6.03	6.8593003	6.4718544	6.3271558	7.1819844	5.4173425
B.2.4.	2.3862313	3.9813189	3.3932875	4.0150841	3.4977707	3.0201987	4.97630385	5.7777937	5.416641	5.2742393	6.0893349	4.3051249
B.2.5.	2.3289697	3.9926057	3.4065232	3.9813189	3.4589594	3.0350618	4.9853385	5.77	5.4249424	5.2827644	6.0671245	4.2736518
B.2.6.	2.4787295	3.9926057	3.4065232	4.0262762	3.4589594	3.0938003	4.9853385	5.7855769	5.4083269	5.2485808	6.0671245	4.2736518
B.2.7.	2.6027101	4.0262762	3.4459251	4.0596675	3.4459251	3.1939944	5.01234476	5.8088639	5.416641	5.24	6.059703	4.2631092
B.2.8.	4.8655524	6.4048731	5.7924865	6.4677972	5.8543061	5.4850798	7.34534546	8.197219	7.7789781	7.6195866	8.4411196	6.6957897
B.2.9.	2.8176054	4.5204978	3.8517009	4.5105321	3.9440588	3.5236912	5.4	6.2660115	5.8753383	5.7628118	6.5154892	4.8310351

	B.4.1.	B.4.2.	B.4.3.	B.4.4.	B.4.5.	B.4.6.	B.4.7.
B.2.1.	9.7846001	9.9571281	11.407121	11.738824	12.584947	13.378625	11.5277101
B.2.2.	10.638139	10.801741	12.240507	12.555493	13.4206	14.196281	12.3811995
B.2.3.	7.9882163	8.1531834	9.6116388	9.9355725	10.797875	11.573694	9.71043253
B.2.4.	6.9364905	7.1026826	8.5458118	8.8558512	9.7217488	10.501548	8.66685641
B.2.5.	6.9429749	7.1090154	8.5405445	8.8405939	9.7078525	10.488684	8.67204705
B.2.6.	6.93	7.0963441	8.5405445	8.8405939	9.7078525	10.488684	8.64085644
B.2.7.	6.9364905	7.1026826	8.5458118	8.8355022	9.703216	10.484393	8.62521884
B.2.8.	9.3090494	9.4638047	10.921282	11.224331	12.104582	12.859627	10.994585
B.2.9.	7.4203841	7.5619111	9.0050042	9.3135868	10.215165	10.942948	9.14420035

	B.4.1.	B.4.2.	B.4.3.	B.4.4.	B.4.5.	B.4.6.	B.4.7.

B.3.1.	4.8516389	4.9984398	6.3908059	6.7082039	7.5895257	8.3323226	6.61725774
B.3.2.	2.9902508	3.1913007	4.618831	4.9503131	5.772703	6.5969993	4.74578761
B.3.3.	3.5752622	3.7221096	5.1674462	5.4958985	6.385139	7.1270541	5.32518544
B.3.4.	3.0350618	3.2053705	4.5895098	4.8862665	5.7492695	6.5353194	4.77414914
B.3.5.	3.6252586	3.7701591	5.1674462	5.43	6.3285069	7.0763621	5.29127584
B.3.6.	4.070737	4.2375583	5.6581711	6.0093677	6.8591836	7.6328304	5.85539068
B.3.7.	2.1000238	2.1914607	3.6274095	3.9732732	4.8987345	5.5749529	3.83567465
B.3.8.	1.2351518	1.4185909	2.7924183	3.127555	3.9784545	4.7657528	3.00374766
B.3.9.	1.5591344	1.69	3.1443441	3.4753561	4.3736026	5.1031853	3.28913362
B.3.10.	1.7164207	1.8980253	3.3307807	3.6324785	4.4903897	5.2786457	3.39328749
B.3.11.	1.2941793	1.3145722	2.5828085	2.8259689	3.7884561	4.45	2.77027074
B.3.12.	2.8175166	3.0173498	4.3820543	4.6389654	5.47	6.3045698	4.4708053

Dunn index and the matrix of vertexes of the pyramids

$$I_{dunn} = \frac{1.2351518}{9.10439454} = 0.1356655$$

9,104395	1,3284954	6,423091	11,05445
1,328495	3,8612045	2,32897	6,93
6,423091	2,3289697	3,986741	1,235152
11,05445	6,93	1,235152	3,633235

Dunn index and the matrix of the primary clusters, which are bases of the pyramids

$$I_{dunn} = \frac{0.4534310}{34.9064410} = 0.0129899$$

34,9064410	0,5300000	5,2047000	10,4500000
0,5300000	5,0707100	0,4534310	5,4613550
5,2047000	0,4534310	5,0114770	0,4609770
10,4500000	5,4613550	0,4609770	4,1868840

Table 6. Case C. Five clusters based on the Chi-merge discretization after the final pruning

	Class	N	P	K	Crop	variety
C.1.1	1	0	0	0	0	-

	Class	N	P	K	Crop	variety
C.2.1.	2	1.2N	0P	0.6K	27.58	Kharkov 81
C.2.2.	2	0N	0P	0K	27.69	Mironov 808

	Class	N	P	K	Crop	variety
C.3.1.	3	0.6N	0.6P	0.9K	35.32	Kharkov 81
C.3.2.	3	0.6N	0.9P	0.3K	33.77	Kharkov 81
C.3.3.	3	0.9N	0.6P	0.3K	35.93	Kharkov 81
C.3.4.	3	0.9N	0.6P	0.6K	34.36	Kharkov 81
C.3.5.	3	0N	0P	0K	36.17	Caucasus
C.3.6.	3	0N	0.6P	0.3K	35.29	Caucasus
C.3.7.	3	1.2N	0P	0.6K	36.61	Caucasus
C.3.8.	3	0N	0P	0K	35.73	Mironov jubilee

	Class	N	P	K	Crop	variety
C.4.1.	4	0N	1.2P	0.6K	41.17	Caucasus
C.4.2.	4	0.6N	0P	0.3K	37.93	Caucasus
C.4.3.	4	0.6N	0.6P	0K	42.93	Caucasus
C.4.4.	4	0.6N	0.6P	0.3K	38.96	Caucasus
C.4.5.	4	0.6N	0.6P	0.6K	42.34	Caucasus
C.4.6.	4	0.6N	0.9P	0.3K	37.93	Caucasus
C.4.7.	4	0.6N	1.2P	0.3K	42.93	Caucasus
C.4.8.	4	0.9N	0.6P	0.3K	38.96	Caucasus
C.4.9.	4	1.2N	0.6P	0.3K	38.96	Caucasus
C.4.10.	4	1.2N	1.2P	0.6K	42.34	Caucasus
C.4.11.	4	0.6N	0.6P	0.9K	42.20	Mironov 808
C.4.12.	4	0.6N	0.6P	0.9K	39.89	Mironov jubilee
C.4.13.	4	1.5N	0.6P	0.3K	41.05	Mironov jubilee

	Class	N	P	K	Crop	variety
C.3.1.	5	0.9N	0.6P	0.3K	46.00	Mironov 808
C.3.2.	5	0.9N	0.6P	0.6K	46.05	Mironov 808
C.3.3.	5	0.9N	0.9P	0.6K	47.49	Mironov 808
C.3.4.	5	1.2N	0.6P	0.3K	48.64	Mironov 808
C.3.5.	5	1.2N	1.2P	0K	48.64	Mironov 808
C.3.6.	5	1.2N	1.2P	0.6K	47.77	Mironov 808
C.3.7.	5	1.2N	1.2P	0.9K	49.41	Mironov 808
C.3.8.	5	1.5N	0.6P	0.3K	47.58	Mironov 808

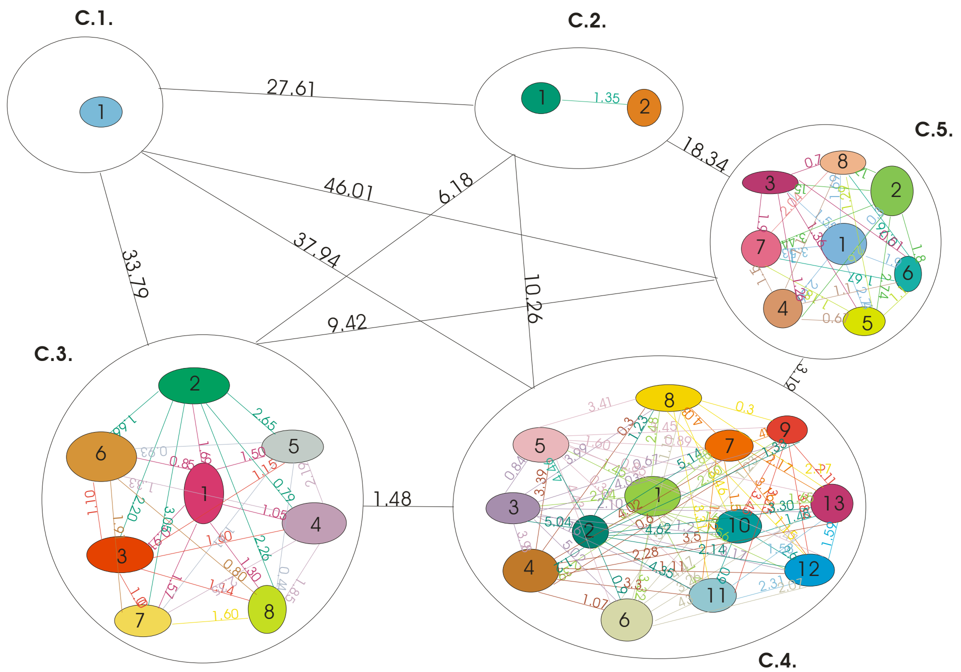


Table 7. Case C. Internal distances and diameters

	C.1.1.
C.1.1.	0

	C.2.1.	C.2.2.
C.2.1.	0	1,346142637
C.2.2.		0

	C.3.1.	C.3.2.	C.3.3.	C.3.4.	C.3.5.	C.3.6.	C.3.7.	C.3.8.
C.3.1.	0	1.688934575	0.9066973	1.04957134	1.5008331	0.8490583	1.57292721	1.30311166
C.3.2.		0	2.20127236	0.78619336	2.64952826	1.66144515	3.05378454	2.25867218
C.3.3.			0	1.59840546	1.14786759	1.10435502	1.00119928	1.14017543
C.3.4.				0	2.19228192	1.32849539	2.34787138	1.84577897
C.3.5.					0	0.92973114	1.41194901	0.44
C.3.6.						0	1.90588562	0.80224684
C.3.7.							0	1.60449369
C.3.8.								0

	C.4.1.	C.4.2.	C.4.3.	C.4.4.	C.4.5.	C.4.6.	C.4.7.	C.4.8.	C.4.9.	C.4.10.	C.4.11.	C.4.12.	C.4.13.
C.4.1.	0	3.51960225	2.04391781	2.3862	1.4453	3.32228837	1.88350737	2.47872951	2.60271013	1.67597733	1.36780847	1.5647364	1.64754363
C.4.2.		0	5.0447993	1.1920	4.46073	0.9	5.14198405	1.22918672	1.33450365	4.61931813	4.35349285	2.13579025	3.3021811
C.4.3.			0	3.98138	0.8415	5.01796772	0.67082039	3.99260567	4.0262762	1.19503138	1.15883562	3.17042584	2.10580151
C.4.4.				0	3.3933	1.07280007	4.01508406	0.3	0.6	3.49777072	3.29508725	1.106752	2.27554389
C.4.5.					0	4.43036116	0.89336443	3.40652315	3.44592513	0.84852814	0.33105891	2.46829901	1.60128074
C.4.6.						0	5.00899191	1.11395691	1.22918672	4.4708053	4.32237203	2.07161773	3.26104278
C.4.7.							0	4.0262762	4.05966747	1.49602807	1.11933016	3.15620025	2.16896289
C.4.8.								0	0.3	3.45895938	3.30871576	1.14669089	2.17441946
C.4.9.									0	3.44592513	3.34926858	3.8440875	2.11142132
C.4.10.										0	0.9108238	2.61007663	1.48462116
C.4.11.											0	2.31	1.57876534
C.4.12.												0	1.58606431
C.4.13.													0

	C.5.1.	C.5.2.	C.5.3.	C.5.4.	C.5.5.	C.5.6.	C.5.7.	C.5.8.
C.5.1.	0	0.30413813	1.54922561	2.65699078	2.74036494	1.91648115	3.52676906	1.69008876
C.5.2.		0	1.47091808	2.624519	2.74191539	1.84618526	3.43941856	1.6705987
C.5.3.			0	1.26194295	1.36473441	0.5083306	1.98907013	0.74033776
C.5.4.				0	0.67082039	1.09859001	1.14581848	1.10163515
C.5.5.					0	1.0568349	1.1844408	1.28980619
C.5.6.						0	1.66721324	0.75901252
C.5.7.							0	2.03933813
C.5.8.								0

Table 8. Case C. External distances between clusters

	C.2.1.	C.2.2.
C.1.1.	27.6126131	27.69

	C.3.1.	C.3.2.	C.3.3.	C.3.4.	C.3.5.	C.3.6.	C.3.7.	C.3.8.
C.1.1.	35.3416525	33.78865046	35.9475298	34.3822571	36.17	35.2963752	36.6345752	35.73

	C.4.1.	C.4.2.	C.4.3.	C.4.4.	C.4.5.	C.4.6.	C.4.7.	C.4.8.	C.4.9.	C.4.10.	C.4.11.	C.4.12.	C.4.13.
C.1.1.	41.1919	37.93593152	42.9384	38.97039	42.3528	37.9466059	42.95200694	38.9761671	38.9842481	42.3782444	42.2181241	39.9091731	41.0828736

	C.5.1.	C.5.2.	C.5.3.	C.5.4.	C.5.5.	C.5.6.	C.5.7.	C.5.8.
C.1.1.	46.0136936	46.06660938	47.5108419	48.6594246	48.6695963	47.8039005	49.44732652	47.6083648

	C.3.1.	C.3.2.	C.3.3.	C.3.4.	C.3.5.	C.3.6.	C.3.7.	C.3.8.
C.2.1.	7.7921499	6.290953823	8.38227296	6.81310502	8.69414171	7.83160903	9.03	8.25969128
C.2.2.	7.72961189	6.182750197	8.31610486	6.78372317	8.48	7.62954782	9.020332588	8.04

	C.4.1.	C.4.2.	C.4.3.	C.4.4.	C.4.5.	C.4.6.	C.4.7.	C.4.8.	C.4.9.	C.4.10.	C.4.11.	C.4.12.	C.4.13.
C.2.1.	13.6955504	10.37171635	15.3851	11.4155	14.7843701	10.41069	15.41144056	11.4037	11.3997544	14.8087	14.6476756	12.3428562	13.4900297
C.2.2.	13.546601	10.26194913	15.26368	11.3059	14.6868138	10.3013	15.30188224	11.3258	11.3535413	14.7602	14.5626268	12.2625446	13.4606686

	C.5.1.	C.5.2.	C.5.3.	C.5.4.	C.5.5.	C.5.6.	C.5.7.	C.5.8.
C.2.1.	18.4346522	18.4821779	19.9325889	21.0706811	21.1026918	20.2256298	21.86501544	20.0134954
C.2.2.	18.3443752	18.40161949	19.849937	20.9950589	21.0186227	20.1605159	21.80477929	19.9577579

	C.4.1.	C.4.2.	C.4.3.	C.4.4.	C.4.5.	C.4.6.	C.4.7.	C.4.8.	C.4.9.	C.4.10.	C.4.11.	C.4.12.	C.4.13.
C.3.1.	5.9188259	2.744467161	7.66303465	3.68911914	7.02640733	2.69482838	7.657160048	3.7013	3.73759281	7.07745717	6.88	4.57	5.8312
C.3.2.	7.43639698	4.256242474	9.16982006	5.19866329	8.58049532	4.16	9.164911347	5.2073	5.23317303	8.60144755	8.4566	6.1567	7.3416
C.3.3.	5.35888048	2.109502311	7.01284536	3.04481527	6.42402522	2.04450483	7.032069397	3.03	3.04481527	6.45198419	6.3058	4.0164	5.1550
C.3.4.	6.89536801	3.644845676	8.59621428	4.61952378	7.98563711	3.60761694	8.601447553	4.6098	4.61952378	8.00814585	7.8515	5.5463	6.7235
C.3.5.	5.17687164	1.883507367	6.81304631	2.93156955	6.25690818	2.08748653	6.898376621	3.0073	3.11032153	6.42720001	6.1556	3.9203	5.1492
C.3.6.	5.9181416	2.773012802	7.66939372	3.7187229	7.08184298	2.72389427	7.686975998	3.7787	3.86120448	7.18279194	6.9619	4.6776	5.9521
C.3.7.	4.86555238	1.480675522	6.40487314	2.51644591	5.79248651	1.73274349	6.467797152	2.4622	2.44386988	5.85430611	5.6620	3.4012	4.5004
C.3.8.	5.6029992	2.3	7.24982758	3.35304339	6.69119571	2.46981781	7.330075034	3.4195	3.51039884	6.85070069	6.5872	4.34	5.5680

	C.5.1.	C.5.2.	C.5.3.	C.5.4.	C.5.5.	C.5.6.	C.5.7.	C.5.8.
C.3.1.	10.7010467	10.73838442	12.1810878	13.3469997	13.3773091	12.4824877	14.11552691	12.3076237
C.3.2.	12.2373567	12.29098857	13.7265582	14.8851234	14.8881463	14.0192724	15.66587374	13.8425467
C.3.3.	10.07	10.12444566	11.5677828	12.71354	12.7312254	11.8627821	13.5100111	11.6654404
C.3.4.	11.6438653	11.69	13.1334268	14.2863011	14.3083332	13.426768	15.06792952	13.2370087
C.3.5.	9.89388195	9.9571281	11.4071206	12.545553	12.5849474	11.7388245	13.37862474	11.5277101
C.3.6.	10.7477486	10.8017406	12.2405065	13.4038241	13.4205998	12.5554928	14.1962812	12.3811995
C.3.7.	9.4187101	9.463804732	10.921282	12.0486887	12.1045818	11.2243307	12.85962674	10.994585
C.3.8.	10.3311616	10.39386357	11.8438845	12.9829927	13.0210637	12.1738079	13.81421008	11.9633816

	C.5.1.	C.5.2.	C.5.3.	C.5.4.	C.5.5.	C.5.6.	C.5.7.	C.5.8.
C.4.1.	4.95871959	4.998439757	6.3908059	7.59545259	7.58952568	6.70820393	8.332322605	6.61725774
C.4.2.	8.09783304	8.153183427	9.61163878	10.7435609	10.7978748	9.93557245	11.57369431	9.71043253
C.4.3.	3.09917731	3.191300675	4.61883102	5.74926952	5.77270301	4.95031312	6.596999318	4.74578761
C.4.4.	7.04638915	7.102682592	8.54581184	9.69857722	9.72174881	8.85585117	10.50154751	8.66685641
C.4.5.	3.68450811	3.722109617	5.16744618	6.33561362	6.385139	5.49589847	7.12705409	5.32518544
C.4.6.	8.08114472	8.13660863	9.5694096	10.7309878	10.7351805	9.86740087	11.5152247	9.69651999
C.4.7.	3.14243536	3.205370493	4.58950978	5.77270301	5.74926952	4.88626647	6.535319426	4.77414914
C.4.8.	7.04	7.096344129	8.54054448	9.68464764	9.70785249	8.84059387	10.48868438	8.64085644
C.4.9.	7.04638915	7.102682592	8.54581184	9.68	9.70321596	8.83550225	10.48439316	8.62521884
C.4.10.	3.73304166	3.770159148	5.16744618	6.33561362	6.32850693	5.43	7.07636206	5.29127584
C.4.11.	3.85875628	3.873306081	5.31545859	6.49566009	6.55771302	5.64224246	7.259758949	5.48765888
C.4.12.	6.14671457	6.174593104	7.61774245	8.79104658	8.8369961	7.93122941	9.557740319	7.76570023
C.4.13.	4.98623104	5.044799302	6.48178988	7.59592654	7.62549015	6.76005917	8.408305418	6.53

Dunn index and the matrix of vertexes of the pyramids

$$I_{dunn} = \frac{1.480675522}{5.14198405} = 0.2879580$$

0	27,6126131	33,7886505	37,93593	46,0136936
27,6126131	1,34614264	6,1827502	10,26195	18,3443752
33,7886505	6,1827502	3,05378454	1,480676	9,4187101
37,9359315	6,1827502	1,48067552	5,141984	3,09917731
46,0136936	10,2619491	9,4187101	3,099177	3,52676906

Dunn index and the matrix of the primary clusters, which are bases of the pyramids

$$I_{dunn} = \frac{0.3354100}{23.8800000} = 0.0140456$$

23,8800000	0,7520640	6,6115430	13,1371380	19,2591300
0,7520640	3,8500000	2,4350770	8,9300000	15,0678600
6,6115430	2,4350770	6,2588260	0,6177380	6,4808640
13,1371380	8,9300000	0,6177380	6,1773780	0,3354100
19,2591300	15,0678600	6,4808640	0,3354100	6,3767080

Table 9. Case D. Two clusters based on merged clusters from case C.

	Class 1				
	N	P	K	Crop	variety
D.1.1.	0N	0P	0K	36.17	Caucasus
D.1.2.	0N	0.6P	0.3K	35.29	Caucasus
D.1.3.	1.2N	0P	0.6K	36.61	Caucasus
D.1.4.	0N	0P	0K	36.16	Kharkov 81
D.1.5.	0.6N	0.6P	0.9K	35.32	Kharkov 81
D.1.6.	0.6N	0.9P	0.3K	33.77	Kharkov 81
D.1.7.	0.9N	0.6P	0.3K	35.93	Kharkov 81
D.1.8.	0.9N	0.6P	0.6K	34.36	Kharkov 81
D.1.9.	0N	0P	0K	27.69	Mironov 808
D.1.10.	0N	0.6P	0.3K	25.86	Mironov 808
D.1.11.	0N	1.2P	0.6K	25.95	Mironov 808
D.1.12.	0N	0P	0K	35.73	Mironov jubilee

	Class 2				
	N	P	K	Crop	variety
D.2.1.	0.6N	0.6P	0.3K	38.96	Caucasus
D.2.2.	0.6N	0.6P	0.6K	42.34	Caucasus
D.2.3.	0.6N	0.6P	0K	42.93	Caucasus
D.2.4.	0.6N	0P	0.3K	37.93	Caucasus
D.2.5.	0.6N	1.2P	0.3K	42.93	Caucasus
D.2.6.	0N	1.2P	0.6K	41.17	Caucasus
D.2.7.	1.2N	0.6P	0.3K	38.96	Caucasus
D.2.8.	1.2N	1.2P	0.6K	42.34	Caucasus
D.2.9.	1.2N	0.6P	0.3K	47.33	Kharkov 81
D.2.10.	-	-	-	49.41	Mironov 808
D.2.11.	0.6N	0.6P	0.6K	42.96	Mironov jubilee
D.2.12.	0.6N	1.2P	0.3K	46.20	Mironov jubilee
D.2.13.	1.2N	0.6P	0.3K	44.20	Mironov jubilee
D.2.14.	1.2N	1.2P	0K	43.17	Mironov jubilee
D.2.15.	1.5N	0.6P	0.3K	41.05	Mironov jubilee

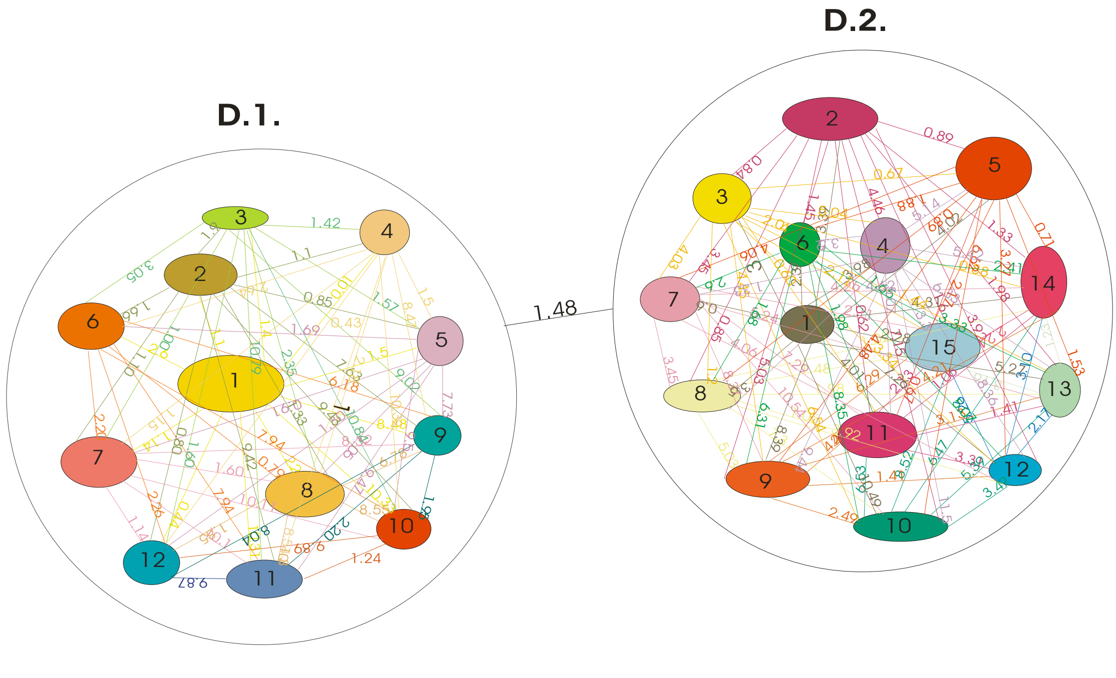


Table 10. Case D. Internal distances and diameters

Classes (1+2+3) = {79}

	D.1.1.	D.1.2.	D.1.3.	D.1.4.	D.1.5.	D.1.6.	D.1.7.	D.1.8.	D.1.9.	D.1.10.	D.1.11.	D.1.12.
D.1.1.	0	1.1065261	1.41194901	0.01	1.5008331	2.64952826	1.14786759	2.19228192	8.48	10.3318004	10.3076865	0.44
D.1.2.		0	1.90588562	1.09859001	0.8490583	1.66144515	1.10435502	1.32849539	7.62954782	9.43	9.42154977	0.80224684
D.1.3.			0	1.41509717	1.57292721	3.05378454	1.00119928	2.34787138	9.02033259	10.8375505	10.7942392	1.60449369
D.1.4.				0	1.4951923	2.64047344	1.14581848	2.18403297	8.47	10.3218215	10.2977716	0.43
D.1.5.					0	1.68893458	0.9066973	1.04957134	7.72961189	9.49797873	9.47031678	1.30311166
D.1.6.						0	2.20127236	0.78619336	6.1827502	7.93839404	7.93992443	2.25867218
D.1.7.							0	1.59840546	8.31610486	10.1101385	10.0965539	1.14017543
D.1.8.								0	6.78372317	8.55277733	8.54272205	1.84577897
D.1.9.									0	1.9490767	2.19718001	8.04
D.1.10.										0	1.2402016	9.89277009
D.1.11.											0	9.87159562
D.1.12.												0

Classes (4+5) = {173}

	D.2.1.	D.2.2.	D.2.3.	D.2.4.	D.2.5.	D.2.6.	D.2.7.	D.2.8.	D.2.9.	D.2.10.	D.2.11.	D.2.12.	D.2.13.	D.2.14.	D.2.15.
D.2.1.	0	3.3933	3.9813	1.1920	4.0151	2.3862	0.6	3.49777072	8.392	10.4886844	4.01123422	7.26481934	5.27423928	4.30512485	2.27554389
D.2.2.		0	0.8415	4.4607	0.8934	1.4453	3.4459	0.84852814	5.035	7.14597089	0.62	3.91785656	1.97727085	1.33	1.60128074
D.2.3.			0	5.0448	0.67082	2.0439	4.0263	1.19503138	4.451	6.53531943	0.60074953	3.33809826	1.43627992	0.88181631	2.10580151
D.2.4.				0	5.1420	3.5196	1.3345	4.61931813	9.438	11.4995826	5.07453446	8.35660816	6.32715576	5.41734252	3.3021811
D.2.5.					0	1.88357	4.0597	0.89336443	4.481	6.62422826	0.67149088	3.27	1.52738338	0.71246053	2.16896289
D.2.6.						0	2.6027	1.67597733	6.312	8.34850885	1.98093412	5.07453446	3.3272962	2.40831892	1.64754363
D.2.7.							0	3.44592513	8.37	10.5400427	4.05585996	7.28955417	5.24	4.26310919	2.11142132
D.2.8.								0	5.035	7.29553973	1.05090437	3.91785656	1.97727085	1.02415819	1.48462116
D.2.9.									0	2.49327094	4.42118762	1.41311712	3.13	4.21373943	6.28716152
D.2.10.										0	6.53318452	3.49200515	5.38832998	6.46665292	8.51995305
D.2.11.											0	3.30871576	1.40982268	1.06023582	2.13262749
D.2.12.												0	2.1725561	3.10336914	5.26236639
D.2.13.													0	1.22918672	3.16425347
D.2.14.														0	2.24374687
D.2.15.															0

Table 11. Case D. External distances between clusters

	D.2.1.	D.2.2.	D.2.3.	D.2.4.	D.2.5.	D.2.6.	D.2.7.	D.2.8.	D.2.9.	D.2.10.	D.2.11.	D.2.12.	D.2.13.	D.2.14.	D.2.15.
D.1.1.	2.932	6.257	6.8131	1.88350737	6.898	5.1769	3.1103	6.42720001	11.244	13.24	6.86906835	10.1237789	8.14683374	7.20277724	5.14921353
D.1.2.	3.719	7.082	7.67	2.7730128	7.687	5.9181	3.8612	7.18279194	12.100	14.14	7.69927919	10.9429475	8.99044493	7.99902494	5.95210887
D.1.3.	2.516	5.792	6.40	1.48067552	6.468	4.8656	2.4439	5.85430611	10.741	12.87	6.40644207	9.68803902	7.6195866	6.69578972	4.50039998
D.1.4.	2.941	6.267	6.826	1.89285499	6.908	5.1865	3.1193	6.43680045	11.254	13.25	6.87895341	10.1336864	8.15669051	7.2124961	5.15869169
D.1.5.	3.689	7.026	7.66	2.74446716	7.657	5.9188	3.7376	7.07745717	12.040	14.14	7.64588778	10.9130381	8.92044842	7.94685472	5.83120056
D.1.6.	5.199	8.580	9.17	4.25624247	9.165	7.4364	5.2332	8.60144755	13.577	15.68	9.19978804	12.4336197	10.4515501	9.42867965	7.34155297
D.1.7.	3.045	6.424	7.01	2.10950231	7.032	5.3589	3.0448	6.45198419	11.404	13.53	7.04279064	10.2918852	8.27543957	7.27719726	5.15503637
D.1.8.	4.620	7.986	8.60	3.64484568	8.601	6.8954	4.6195	8.00814585	12.977	15.10	8.60523097	11.8627821	9.84914209	8.85585117	6.72354817
D.1.9.	11.306	14.687	15.26	10.2619491	15.302	13.547	11.354	14.760166	19.688	21.72	15.3053226	18.5609833	16.5671392	15.5727454	13.4606686
D.1.10.	13.114	16.494	17.08	12.0997893	17.091	15.325	13.155	16.5372428	21.504	23.56	17.1131528	20.3576914	18.3792165	17.3645069	15.2638822
D.1.11.	13.041	16.412	17.01	12.0586235	16.993	15.22	13.082	16.4338705	21.424	23.50	17.0311509	20.2611081	18.3017076	17.2721857	15.1891409
D.1.12.	3.353	6.691	7.25	2.3	7.330	5.603	3.5104	6.85070069	11.681	13.68	7.30430695	10.5598722	8.58084495	7.63109429	5.56797989

Dunn index and the matrix of vertexes of the pyramids

$$I_{dunn} = \frac{1.48067552}{11.4995826} = 0.1287591$$

10,83755	1,4806755
1,480676	11,499583

Dunn index and the matrix of the primary clusters, which are bases of the pyramids

$$I_{dunn} = \frac{0.6177380}{36.6857680} = 0,0168386$$

36,6857680	0,6177380
0,6177380	12,4967240

Authors' Information

***Krassimira Ivanova** – University of National and World Economy, Sofia, Bulgaria*
e-mail: krasy78@mail.bg
Major Fields of Scientific Research: Data Mining

***Vitalii Velychko** – Institute of Cybernetics, NASU, Kiev, Ukraine*
e-mail: Velychko@rambler.ru
Major Fields of Scientific Research: Data Mining, Natural Language Processing

***Krassimir Markov** – Institute of Mathematics and Informatics at BAS, Sofia, Bulgaria;*
e-mail: markov@foibg.com
Major Fields of Scientific Research: Multi-dimensional information systems, Data Mining