

# Analysis of regional development disparities in Ukraine with fuzzy clustering technique

Kateryna Gorbatiuk<sup>1,\*</sup>, Olha Mantalyuk<sup>1</sup>, Oksana Proskurovych<sup>1</sup>, and Oleksandr Valkov<sup>1</sup>

<sup>1</sup>Khmelnytskyi National University, Department of Automated Systems and Modeling in Economics, Khmelnytski, Ukraine

**Abstract.** Disparities in the development of regions in any country affect the entire national economy. Detecting the disparities can help formulate the proper economic policies for each region by taking action against the factors that slow down the economic growth. This study was conducted with the aim of applying clustering methods to analyse regional disparities based on the economic development indicators of the regions of Ukraine. There were considered fuzzy clustering methods, which generalize partition clustering methods by allowing objects to be partially classified into more than one cluster. Fuzzy clustering technique was applied using R packages to the data sets with the statistic indicators concerned to the economic activities in all administrative regions of Ukraine in 2017. Sets of development indicators for different sectors of economic activity, such as industry, agriculture, construction and services, were reviewed and analysed. The study showed that the regional cluster classification results strongly depend on the input development indicators and the clustering technique used for this purpose. Consideration of different partitions into fuzzy clusters opens up new opportunities in developing recommendations on how to differentiate economic policies in order to achieve maximum growth for the regions and the entire country.

## 1 Introduction

Economic policies that take into account differences in regional development should be coordinated using scientific approaches to achieve maximum results in each region and for the whole country. This article is dedicated to the problem of clustering Ukrainian regions in different groups accordingly to their economic development levels. The usefulness of such division is obvious. Really, having at disposal the partitioning into different clusters based on economic indicators, a decision maker can elaborate economic policy measures, which are specific for every cluster and similar for all the regions inside the same cluster. So, the number of policy options substantially reduce in comparison with the case, when the decision is made on each particular region.

Clustering also provides an opportunity to identify groups of regions that are most attractive as objects of domestic and foreign investment. Undoubtedly, the use of cluster analysis for improving regional policy will increase the efficiency of the economic system as a whole, which is especially important for today's Ukraine and is a necessary condition for its economic growth.

Nowadays, a good deal of research representing manifold of cluster analysis approaches and tools has been conducted and reflected at the relevant literary sources. Nevertheless, search for the most acceptable clustering methods still retains its relevance. The reason is that every method has its own advantages and disadvantages.

Fuzzy clustering methods permit the gradual assessment of the membership of data elements in a cluster which is described by a membership function

valued in the real unit interval  $[0; 1]$ . So, in fuzzy clustering it is assumed that the boundaries between groups are not well defined, like in the case of most natural systems. Therefore, fuzzy clustering approaches make it possible to more adequately describe and solve the real problem, such as estimating regional development disparities.

This article presents a study on application of hard cluster analysis methods and clustering methods based on fuzzy sets theory. A new approach to evaluating regional disparities in Ukraine using a fuzzy clustering technique is given. There were used statistical data on indicators of economic activities in different regions of Ukraine in 2017. The considered methods are especially useful for the case of qualitative economic indicators.

This article consists of six sections. The first one substantiates the background of the conducted research. In the second section, review of the scientific literature on research topic is presented. The third part reveals the theoretical basis of the proposed clustering techniques. The course of the study and its main results are presented in the fourth and fifth parts of this paper. The final part contains conclusions based on the research results and discussing areas for the further studies in the field of exploring fuzzy clustering methods and adapting them to regional clustering tasks.

## 2 Literature review

Regional disparities are closely connected to unequal economic development of the regions in different sectors. So, economic disparities are associated with

\* Corresponding author: [kt\\_datsyuk@ukr.net](mailto:kt_datsyuk@ukr.net)

differences in regional qualitative and quantitative economic indicators. Economic disparities are generally assessed using such indicator as gross national product (GNP), combined with the analysis of tax revenues, the growth of industry and agriculture, demographic trend, infrastructure and services [1].

Studies on estimating and classifying of regional development disparities have been performed by many researchers [1-9] all over the world. The most common approaches, which are used for this purpose, are econometric modelling [4, 8, 9], Klassen typology theory [1-3] and different clustering techniques [1, 2, 6]. At the same time, among clustering methods, k-means clustering and hierarchical clustering are most widely used.

The Klassen typology and the developed fuzzy-Klassen model are discussed in the paper [3] along with giving the recommendations on their use in modelling regional development disparities.

The use of clustering techniques in the tasks of classification of regions by the level of economic indicators is represented in articles [1, 2, 6]. Also, there were proposed to join the traditional clustering approaches with fuzzy methods, based on fuzzy sets theory of L. Zadeh [10], and a lot of researches were done to apply them in practice.

The theoretical basics on clustering methods, fuzzy clustering algorithms and their program software implementations are considered in numerous works [11-30]. In this study, we used the fuzzy clustering approaches to identify disparities in the development of Ukrainian regions, which allow us to explore and utilize the advantages of this technique.

### 3 Research methodology

Clustering is one of the important data mining techniques that enable the discovery of hidden relationships from data [15]. The goal of the clustering is to divide the set of data items into several number of groups  $c$ , called clusters. The result of any cluster algorithm is the mapping of data items to a specific group.

In general, clustering techniques are divided into two types, Hierarchical and Partitioned clustering [22]. Partition clustering algorithms divide the data sets into clusters assigning dissimilar data objects to different clusters.

Hierarchical cluster techniques are generally classified into two types, which are agglomerative and divisive clustering [22]. These cluster methods form a dendrogram, which represents nested grouping pattern and similarity level in classification process. At certain group level, dendrogram will break into another group level, thus producing a different data group. In hierarchical clustering, objects that belong to a child cluster also belong to the parent cluster [13].

Hierarchical cluster methods classify data by similarity of distance between two data points. The classical methods for distance measures are Euclidean

and Manhattan distances, which are defined as follow [19]:

$$d_{\text{euc}}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}, \quad (1)$$

$$d_{\text{man}}(x, y) = \sum_{i=1}^n |x_i - y_i|, \quad (2)$$

where  $x$  and  $y$  – two vectors of length  $n$ ;  $d_{\text{euc}}(x, y)$  – Euclidean distance;  $d_{\text{man}}(x, y)$  – Manhattan distance.

Also, there are many other methods to calculate the distance information, but the right choice of distance measures, which depends on the type of the data and the researcher questions, is very important, as it has a strong influence on the clustering results [19].

The conventional (hard or hard) clustering methods restrict that each point of the data set belongs to exactly one cluster [14]. Fuzzy set theory proposed by Zadeh [10] in 1965 gave an idea to describe the uncertainty of belonging to particular class by a membership function. Applications of fuzzy set theory in cluster analysis were early proposed in the work of Bellman, Kalaba, Zadeh [23] and Ruspini [17].

Basic fuzzy clustering techniques include: fuzzy clustering based on fuzzy relation, fuzzy clustering based on objective functions, and the fuzzy generalized K-nearest neighbour rule – one of the powerful nonparametric classifiers [14].

For all fuzzy clustering algorithms, it is necessary to pre- assume the number  $c$  of clusters because, in general, the number  $c$  should be unknown [14]. The quality of the classification of data into partitions depends on the value of the parameter  $c$  that is provided to the algorithm [15].

Fuzzy clustering is a soft clustering technique for classifying data into groups. In fuzzy clustering each data point belongs to all the clusters with varying memberships and these membership values range between zero and one [15].

Most of the clustering algorithms follow a similar structure [11]: (1) select initial cluster centers, (2) calculate the distances between all points and all cluster centers, (3) update the partition matrix until some termination threshold is met. In particular, the classification of fuzzy algorithms is represented in [11].

The most well-known fuzzy clustering algorithms are: fuzzy c-means, fuzzy k-means, (ISODATA), Gustafson Kessel (GK) algorithm [13] etc.

The fuzzy c-means (FCM) algorithm involves the processes in which there is calculation of cluster centers and assignment of points to these centers using a formula of Euclidian distance [13]. The fuzzy c-means algorithm is one of the most widely used fuzzy clustering algorithms. It is a soft clustering algorithm which was firstly studied by Dunn (1973) [28] and generalized by Bezdek (1974; 1981) [29, 30]. The centroid of a cluster is calculated as the mean of all points, weighted by their degree of belonging to the cluster [19]. The above process is kept on repeating itself until the stabilization of cluster centers.

This algorithm assigns a membership value to the data items for the clusters within a range of 0 to 1. Thus, the concepts of fuzzy sets of partial membership are

incorporated and forms overlapping clusters for supporting it [13]. Consequently, the data objects closer to the centers of clusters have higher degrees of membership than objects scattered in the borders of clusters [20].

We can apply clustering algorithms using the R software. The following R packages are used for calculations in our research: 1) *cluster*, *ppclust* and *fclust* for computing fuzzy clustering and 2) *factoextra* for visualizing clusters [27].

The function *hclust()* (*cluster* R package) performs a hard hierarchical cluster analysis using a set of dissimilarities for the *n* objects being clustered. Initially, each object is assigned to its own cluster and then the algorithm proceeds iteratively, at each stage joining the two most similar clusters, continuing until there is just a single cluster. At each stage distances between clusters are recomputed according to the particular clustering method being used [26].

The function *fanny()* (*cluster* R package) can be used to compute fuzzy clustering [26]. It stands for fuzzy analysis clustering and returns an object including the following components: the fuzzy membership matrix containing the degree to which each observation belongs to a given cluster; Dunn’s partition coefficient (a low value indicates a very fuzzy clustering, whereas a value close to 1 indicates a near-hard clustering); the clustering vector containing the nearest hard grouping of observations etc. [19].

The function *fcm()* (*ppclust* R package), which applies the fuzzy *c*-means algorithm also can be used to compute fuzzy clustering. It returns an object including the following components: the fuzzy membership matrix; Initial and final cluster prototypes matrices; the Dunn’s Fuzziness Coefficients; the within cluster sum of squares by cluster etc. [19].

## 4 Case study: grouping regions using different clustering techniques

### 4.1 Data set description

The data for our study was taken from the State Statistics Service of Ukraine [31]. We used the statistic information about the economic activities in 2017 taken by all regions. There we selected some basic indicators of economic activities and we divided them into two groups by their meaning. So, the first group included the indicators of the extraction of aquatic bioresources and the agriculture activities, and the second group included the indicators of the retail trade, services and the industrial activities. All of them were explored and their corresponding values were used in clustering analysis of the regional development. The list of those indicators and their summary statistics are presented in the Tables 1, 2.

In the Table 3, the column “Id” contains the inner identification number of the region which is used for convenience for all following computing results and outputs.

**Table 1.** First group of indicators with their statistics.

Indicator	Mean	Median	St. Dev.	Range
Extraction of aquatic bioresources by fishery water bodies	3793.6	1207.5	5648.1	25163
Value of agricultural products sold by agricultural enterprises	14404.6	14571.7	8143.1	31251
The cost of sold agricultural products from plant growing	11826.2	13293.4	6483.3	20141
The cost of sold agricultural products from animal husbandry	2578.5	2079.9	2672.2	12694
Dynamics of sown areas of agricultural crops, all categories	1158.9	1194.3	520.4	1757.2
Dynamics of sown areas of agricultural crops, agricultural enterprises	811.5	916.1	416.8	1311.5
Dynamics of sown areas of agricultural crops, agricultural households	347.6	307.0	139.3	494.5

**Table 2.** Second group of indicators with their statistics.

Indicator	Mean	Median	St. Dev.	Range
Regional structure of turnover of retail trade	27429.9	19576.8	18943.0	66619.1
Completed construction works	3344.9	1932.9	3114.4	10659.9
Volume of industrial products sold	82165.9	43406.2	95282.4	407469.9
Regional structure of retail trade turnover of retail enterprises	18757.8	12174.2	15042.1	52768.5
Used fuel, (10 <sup>3</sup> )	4075.6	1961.8	5092.7	18645.3
Volume of services sold by enterprises in the service sector, (10 <sup>3</sup> )	3994.8	2108.6	4053.2	14932.1

We considered the values of these indicators, gathered in 2017, for all 24 administrative regions in Ukraine (Table 3).

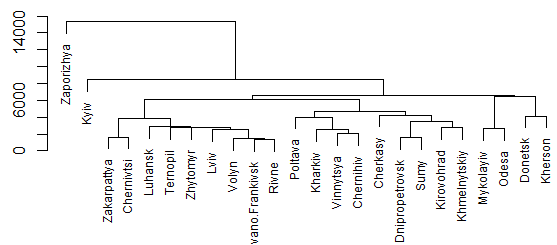
**Table 3.** Administrative regions of Ukraine.

Id	Region	Id	Region
1	Vinnytsya	13	Mykolayiv
2	Volyn	14	Odesa
3	Dnipropetrovsk	15	Poltava
4	Donetsk	16	Rivne
5	Zhytomyr	17	Sumy
6	Zakarpatya	18	Ternopil
7	Zaporizhya	19	Kharkiv
8	Ivano-Frankivsk	20	Kherson
9	Kyiv	21	Khmelnyskiy
10	Kirovohrad	22	Cherkasy
11	Luhansk	23	Chernivtsi
12	Lviv	24	Chernihiv

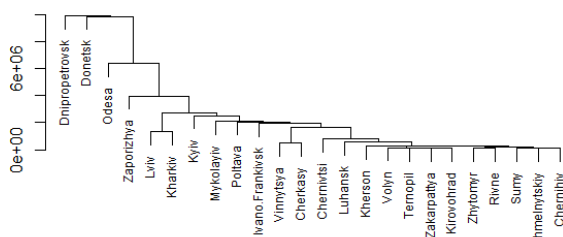
So, there were built two data sets accordingly to each set of indicators. We denoted them as the First data set and the Second Data set. Then, we used both data sets for clustering the regions, based on different groups of indicators, and compared the results.

### 4.2 Clustering results

Before starting the fuzzy clustering analysis, we can apply the hierarchical clustering method, using a linkage method “single”, to both data sets. The results of clustering are illustrated by the cluster dendrograms (Fig. 1, 2), where we can see the data points hierarchically arranged into larger groups dependently on the distances between them.



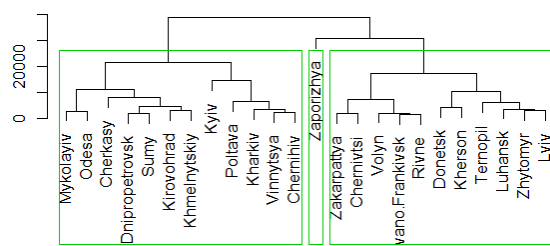
**Fig. 1.** Results of hierarchical clustering for the First data set.



**Fig. 2.** Results of hierarchical clustering for the Second data set.

#### 4.2.1 Three clusters

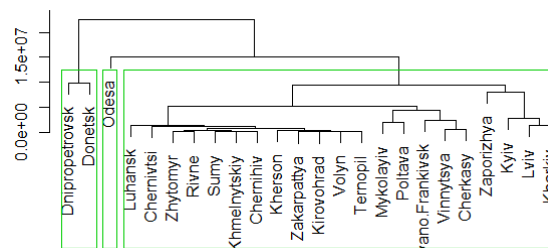
For the number of clusters equal to three ( $c = 3$ ) we conducted the hierarchical clustering by *hclust()* function [26], using a linkage method “complete”, and obtain the hard clusters for two data sets (Fig. 3-4, Table. 4).



**Fig. 3.** Three clusters for the First data set.

The fuzzy clustering methods, applied to both data sets, allowed us to obtain the fuzzy clusters which are characterized by membership coefficients indicated the strength of belonging to the particular cluster for all regions.

We illustrated the fuzzy clusters by several charts (Fig. 5-6) and the table with the values of the membership coefficients obtained by the *fcm()* function [19] (Table 5). The values of membership coefficients vary from 0 to 1 and indicate with different conditional formatting pattern the strength of belonging to the particular cluster for all regions.



**Fig. 4.** Three clusters for the Second data set.

**Table 4.** Hierarchical clustering results for  $c = 3$ .

Clusters	Regions	
	First data set	Second data set
1	1, 3, 9, 10, 13, 14, 15, 17, 19, 21, 22, 24	3, 4
2	7	14
3	2, 4, 5, 6, 8, 11, 12, 16, 18, 20, 23	1, 2, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24

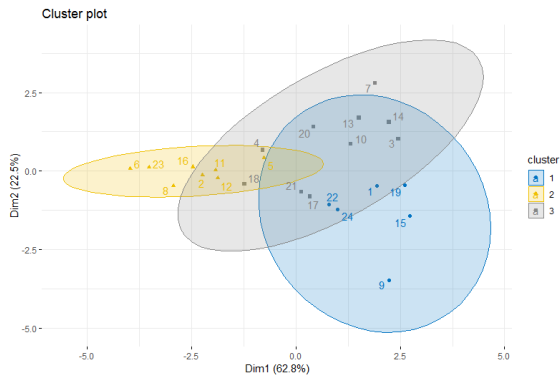
**Table 5.** Membership coefficients for three clusters.

Region	Cluster 1	Cluster 2	Cluster 3
1	0.0069988	0.9614778	0.0315234
2	0.9883655	0.0037207	0.0079138
3	0.0566142	0.3906423	0.5527436
4	0.4711571	0.0926725	0.4361704
5	0.8075004	0.0546427	0.1378569
6	0.8708132	0.0448172	0.0843696
7	0.2282767	0.2399224	0.5318009
8	0.9796605	0.0066517	0.0136878
9	0.0922669	0.6869528	0.2207803
10	0.1287994	0.1990442	0.6721564
11	0.961364	0.0117415	0.0268945
12	0.9448046	0.0166844	0.038511
13	0.0256044	0.0501941	0.9242015
14	0.0454997	0.1413717	0.8131286
15	0.0240021	0.9032314	0.0727665
16	0.9895505	0.003367	0.0070825
17	0.0661397	0.4907331	0.4431272
18	0.5549239	0.1266133	0.3184627
19	0.0109116	0.9509052	0.0381831
20	0.1418759	0.0755601	0.782564
21	0.1426764	0.3478379	0.5094857
22	0.0523409	0.5193732	0.4282859
23	0.9098461	0.0304336	0.0597202
24	0.0167552	0.9196494	0.0635953

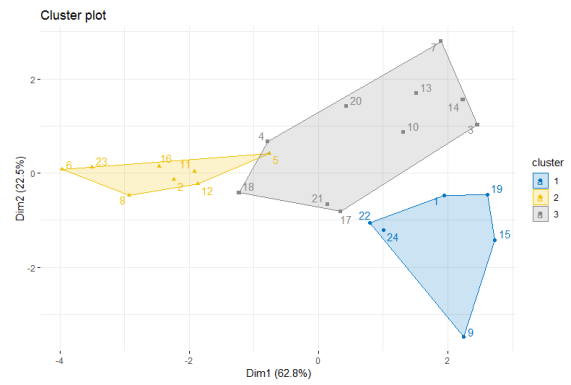
The next plot (Fig. 5) shows the overlapping clusters on the set of all data points. It is the scatterplot of the first two principal components which were derived from the data. It also says that, in our case, 85.3% (62.8%+22.5%) of the information about the multivariate data is captured by this plot.

On the following plot (Fig. 5, 6), the data points with the highest values of the membership coefficients are combined into three different clusters to determine which data points more likely are in each cluster.

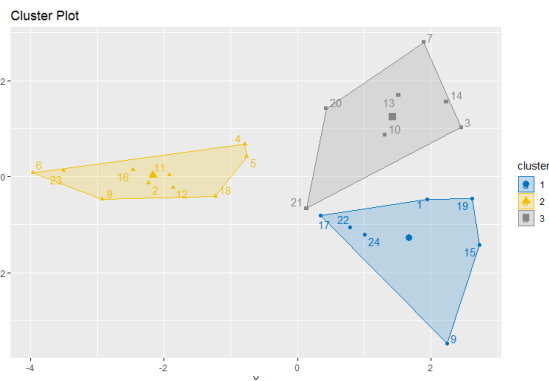




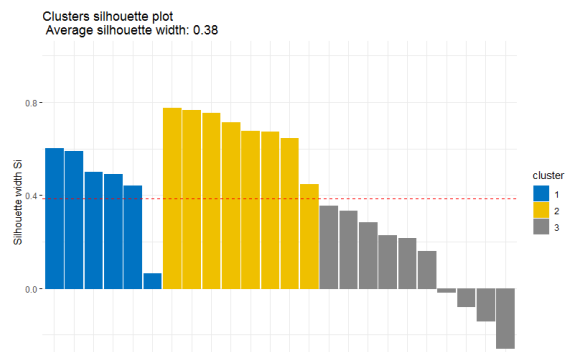
**Fig. 5.** Plot of three fuzzy clusters for the First data set.



**Fig. 8.** Plot of three combined clusters by *fanny()* function for the First data set.

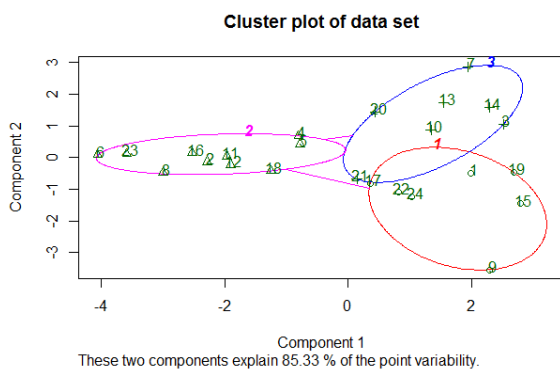


**Fig. 6.** Plot of three combined clusters for the First data set.



**Fig. 9.** Plot of the silhouette coefficients for the First data set.

The similar information is shown on the scatterplot (Fig. 7), which says that 85.33 % of the information about the multivariate data is explained by two principal components.



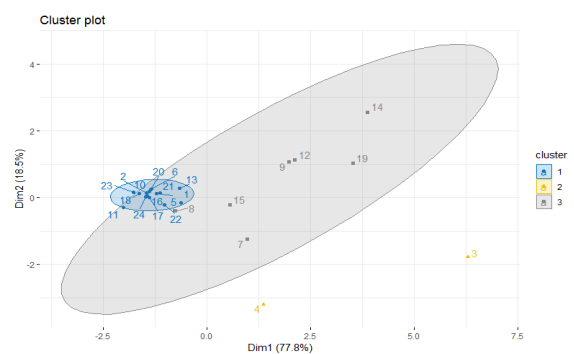
**Fig. 7.** Plot of the fuzzy clusters for the First data set.

Another fuzzy clustering method *fanny()* [26] gave us a slightly different result (Fig. 8).

To estimate the goodness of the clustering results, we can plot the silhouette coefficients which quantify the quality of clustering achieved. The silhouette plot (Fig. 9) displays a measure of how close each point in one cluster is to points in the neighbouring clusters and allows to determine the optimal number of clusters visually.

The plot of silhouette coefficients, built by the last clustering results, shows the average level of the silhouette width 0.38. It is not sufficient result and we can see that some data points are not enough close to points in the neighbouring clusters. Especially, the points in the third cluster are very close to the decision boundary between two neighbouring clusters or even might have been assigned to the wrong cluster.

A similar analysis was performed for the Second data set (Fig. 10-11).

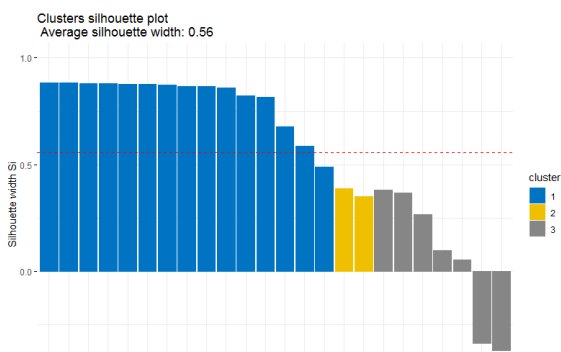


**Fig. 10.** Plot of the fuzzy clusters for the Second data set.

The scatterplot of two principal components (Fig. 10), which were derived from the data, shows the overlapping clusters on the set of all data points, and also, we can see that around 96.3% (77.8%+18.5%) of the information about the multivariate data is explained by these components.

Then, the data points with the highest values of the membership coefficients combined into three different clusters are presented in the Table 6 and show which of them more likely are in each cluster.

The plot of silhouette coefficients (Fig. 11), built by the clustering results of *fanny()* method applied to the Second data set, shows the average level of the silhouette width 0.56. It is rather sufficient result and we can see that most of data points are assigned to the right cluster. But some of them are still on the wrong place.



**Fig. 11.** Plot of the silhouette coefficients for the Second data set.

The summarized results of fuzzy clustering by *fcm()* function applied to both data sets are presented in the Table 6.

**Table 6.** Fuzzy clustering results for  $c = 3$ .

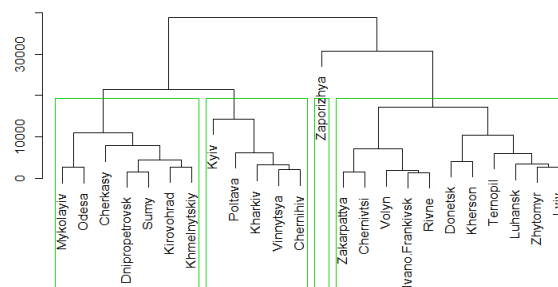
Clusters	Regions	
	First data set	Second data set
1	2, 4, 5, 6, 8, 11, 12, 16, 18, 23	7, 9, 12, 14, 19
2	3, 7, 10, 13, 14, 20, 21	1, 2, 5, 6, 8, 10, 11, 13, 15, 16, 17, 18, 20, 21, 22, 23, 24
3	1, 9, 15, 17, 19, 22, 24	4, 3

As we can see, there were obtained the three fuzzy clusters for each set of economic indicators, and the different partitions of Ukrainian regions show the regional development disparities, which could be analysed and used in decision making process concerned to the economic strategies.

Looking at the fuzziness of these partitions, we can admit that the regions with the average values of membership coefficients are on the boundary of the neighbour clusters, and the strategies for them must be the mixture of the corresponding strategies of the neighbour clusters.

#### 4.2.2 Four clusters

The similar clustering analysis (Fig. 12) were conducted for the case of four clusters ( $c = 4$ ). The results obtained by hierarchical clustering (*hclust()*, “complete”) are in the Table 7.

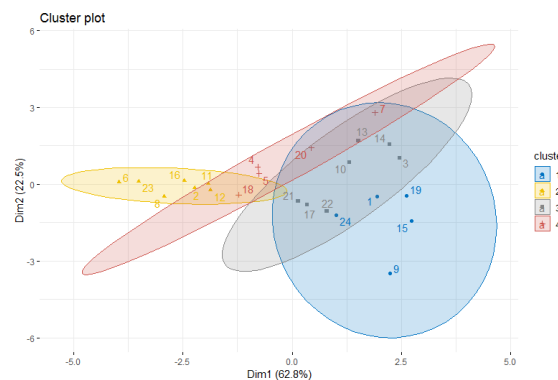


**Fig. 12.** Four clusters for the First data set.

**Table 7.** Hierarchical clustering results for  $c = 4$ .

Clusters	Regions	
	First data set	Second data set
1	3, 10, 13, 14, 17, 21, 22	3
2	1, 9, 15, 19, 24	4
3	7	14
4	2, 4, 5, 6, 8, 11, 12, 16, 18, 20, 23	1, 2, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24

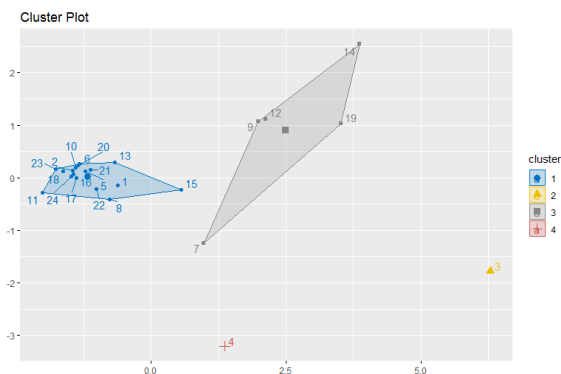
The fuzzy clusters also are presented by different values of membership coefficients (we do not place them here because of the size). But these fuzzy clusters are quite completely described by the overlapping shapes at Fig. 13 and we can say that the plot of two principal components capture around 85.3% of the information about the multivariate data.



**Fig. 13.** Plot of the fuzzy clusters for the First data set.

The fuzzy clusters based on the Second data set we represented by the plot, where the data points with the highest values of the membership coefficients are combined into four different clusters (Fig. 14). Here we have the only two big groups of data points and two data points are stand alone in different clusters. So, the further analysis with larger number of clusters is not rational.

The summarized results of fuzzy clustering by *fcm()* function applied to both data sets are presented in the Table 8.



**Fig. 14.** Plot of the fuzzy clusters for the Second data set.

**Table 8.** Fuzzy clustering results for  $c = 4$ .

Clusters	Regions	
	First data set	Second data set
1	1, 9, 15, 19, 24	1, 2, 5, 6, 8, 10, 11, 13, 15, 16, 17, 18, 20, 21, 22, 23, 24
2	7, 13, 14, 20	3
3	3, 10, 17, 18, 21, 22	7, 9, 12, 14, 19
4	2, 4, 5, 6, 8, 11, 12, 16, 23	4

So, there were obtained the four fuzzy clusters for each set of economic indicators. These different classifications of Ukrainian regions show the disparities in regional development, which can be analysed and used in the decision-making process concerning economic strategies. Including into the analysis the fuzzy nature of obtained partitions, we will gain the new quality of forming of the economic strategies for different regions.

## 5 Results and discussion

The results of fuzzy clustering obtained in this study allows to consider in more detail the similarities in the economic development levels of the Ukrainian regions, which are assigned to the same clusters, and reveal the dissimilarities between the regions assigned to the different clusters. The membership coefficients give us the information how far are the development levels within clusters and between clusters.

This alternative approach can help determine the regional development disparities according to certain indicators. As we showed in this research, the results of partitioning strongly depend on the indicators selected for the analysis, and any clustering technique should be used only along with the substantial analysis of the subject of interest. Before conducting fuzzy clustering, in order to ensure proper economic interpretation of clustering results, a profound analysis of the nature of all economic indicators and relationships between them should be used.

In general, fuzzy clustering results could not be significantly different from hard clustering results. It is quite reasonable, and we could see this in practice. Although the concepts of hard and fuzzy clustering are rather different, they have common features, and the

clusters obtained by different methods predominantly overlap.

The main findings in this research were the conclusions about the regional disparities in the levels of different kinds of economic activities in Ukraine in 2017. Thus, after the analysis of most agricultural indicators, we mark that among Ukrainian regions, Zaporizhyya is the region, which level is significantly different from others. But the analysis of most industrial indicators allows to sign that Dnipropetrovsk and Donetsk regions, as well, are the regions, which levels significantly differ from others.

## 6 Conclusion

Regional disparities in economic development level had been analysed in this study by different clustering techniques. We obtained the classifications based on two groups of economic indicators observed in 2017 for all Ukrainian regions. Now, we can conclude that the regional inequalities across Ukrainian regions can be reduced by the right economic policies if the information about the actual magnitude of differences between the regions will be available before the decision-making process. The fuzzy clustering methods give us the instrument for the estimating these degrees of differences based on the analysis of regional economic activities in target sectors.

We showed, that implementation of fuzzy clustering methods in analysis of regional disparities have many advantages, but it needs to be accompanied with the cluster validity process and substantial analysis of the economic indicators, which we take as the base of the clustering investigation. In further researches, we need to take into consideration the necessity of aggregating the different fuzzy clustering results for developing recommendations on how to differentiate economic policies in order to achieve maximum growth for the regions and the entire country.

## References

1. Munandar, Tb.: Analysis of Regional Development Disparity with Clustering Technique Based Perspective. *International journal of advanced research in computer science*. **6**(1), 137–141 (2015)
2. Munandar, Tb., Azhari, Musdholifah, A., Arsyad, L.: Modified agglomerative clustering with location quotient for identification of regional potential sector. *Journal of theoretical and applied information technology*. **95**(5), 1191–1199 (2017)
3. Munandar, Tb., Retantyo, W.: Fuzzy-Klassen Model for Development Disparities Analysis based on Gross Regional Domestic Product Sector of a Region. *International Journal of Computer Applications*. **123**(7), 17–22 (2015)
4. Novkovska, B.: Regional Development Disparities And Their Connection With Hidden Economy. *UTMS Journal of Economics*. **8**(2), 151–158 (2017)

5. Lukovics, M.: Measuring Regional Disparities on Competitiveness Basis. In: Bajmócy, Z., Lengyel, I. (eds.) *Regional Competitiveness, Innovation and Environment*, pp. 39–53. JATEPress, Szeged (2009)
6. Nosova, O.: The Innovation Development in Ukraine: Problems and Development Perspectives. *International journal of innovation and business strategy*. **2**, 1–13 (2013)
7. Spicka, J.: The Economic Disparity in European Agriculture in the Context of the Recent EU Enlargements. *Journal of Economics and Sustainable Development*. **4**(15), 125–134 (2013)
8. Hryhorkiv, V., Verstiak, A., Verstiak, O., Hryhorkiv, M.: Regional Economic Growth Disparities in Ukraine: Input-Output Analysis Approach. *Scientific Annals of Economics and Business*. **64**(4), 447–457 (2017)
9. Maksymenko, S.: Ukraine's Regional Economic Growth and Analysis of Regional Disparities. Working Paper 6053, Department of Economics, University of Pittsburgh (2016)
10. Zadeh, L.A.: Fuzzy Sets. *Information and Control*. **8**, 338–358 (1965).
11. Gosain, A., Dahiya, S.: Performance Analysis of Various Fuzzy Clustering Algorithms: A Review. *Procedia Computer Science*. **79**, 100–111 (2016)
12. Miller, D.J., Nelson, C., Cannon, M.B., Cannon, K.P.: Comparison of Fuzzy Clustering Methods and Their Applications to Geophysics Data. *Applied Computational Intelligence and Soft Computing*. **2009**, Article 876361 (2009)
13. Malhotra, V.K., Kaur, H., Alam, M.A.: An Analysis of Fuzzy Clustering Methods. *International Journal of Computer Applications*. **94**(19), 9–12 (2014)
14. Yang, M.-S.: A survey of fuzzy clustering. *Mathematical and Computer Modelling*. **18**(11), 1–16 (1993)
15. Joopudi, S., Rathi, S.S., Narasimhana, S., Rengaswamyb, R.: A New Cluster Validity Index for Fuzzy Clustering. *IFAC Proceedings Volumes*. **46**(32), 325–330 (2013)
16. Ruspini, E.H.: Numerical methods for fuzzy clustering. *Information Sciences*. **2**(3), 319–350 (1970)
17. Ruspini, E.H.: A new approach to clustering. *Information and Control*. **15**(1), 22–32 (1969).
18. Gitman, I., Levine, M.D.: An Algorithm for Detecting Unimodal Fuzzy Sets and its Application as a Clustering Technique. *IEEE Transactions on Computers*. **C-19**(7), 583–593 (1970)
19. Kassambara, A.: Advanced clustering. DataNovia. <http://www.datanovia.com/en/courses/advanced-clustering> (2018). Accessed 22 Jan 2019
20. RPubS, Unsupervised Learning - Clustering Fuzzy C Means. <http://rpubs.com/rahulSaha/Fuzzy-CMeansClustering> (2018). Accessed 12 Feb 2019
21. Ferraro, M.B., Giordani, P.: A toolbox for fuzzy clustering using the R programming language. Workshop on Clustering methods and their applications. [http://pro1.unibz.it/projects/Clustering\\_Methods\\_2014/Ferraro.pdf](http://pro1.unibz.it/projects/Clustering_Methods_2014/Ferraro.pdf) (2014). Accessed 18 Jan 2019
22. Jain, A.K., Murty, M.N., Flynn, P.J.: Data Clustering: A Review. *ACM Computing Surveys*. **31**(3), 264–323 (1999)
23. Bellman, R., Kalaba, R., Zadeh, L.A.: Abstraction and pattern classification. *Journal of Mathematical Analysis and Applications*. **13**(1), 1–7 (1966).
24. Giordani, P., Ferraro, M.B., Serafini, A.: Package 'fclust'. CRAN. <http://cran.r-project.org/web/packages/fclust/fclust.pdf> (2018). Accessed 12 Feb 2019
25. Cebeci, Z., Yildiz, F., Kavlak, A.T., Cebeci, C., Onder, H.: Package 'pplust'. CRAN. <https://cran.r-project.org/web/packages/ppclust/ppclust.pdf> (2019). Accessed 12 Feb 2019
26. Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M.: Package 'cluster'. CRAN. <https://cran.r-project.org/web/packages/cluster/cluster.pdf> (2018). Accessed 12 Feb 2019
27. Kassambara, A., Mundt, F.: Package 'factoextra'. CRAN. <https://cran.r-project.org/web/packages/factoextra/factoextra.pdf> (2017). Accessed 12 Feb 2019
28. Dunn, J.C.: A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*. **3**(3), 32–57 (1973).
29. Bezdek, J.C.: Cluster validity with fuzzy sets. *Journal of Cybernetics*. **3**(3), 58–73 (1974).
30. Bezdek, J.C.: Pattern recognition with fuzzy objective function algorithms. Kluwer Academic Publishers, Norwell (1981).
31. State Statistics Service of Ukraine. <http://www.ukrstat.gov.ua/> (2019). Accessed 20 Feb 2019