

**ODESSA NATIONAL ACADEMY
OF FOOD TECHNOLOGIES**

**IX ANNUAL SCIENTIFIC CONFERENCE
INFORMATION TECHNOLOGY AND
AUTOMATION – 2016**

Proceedings

Odessa
October 11-14
2016

CONTENTS

АЛЕКСАШИН О.В., ШЕВЧЕНКО К.Л., ГОРКУН В.В., НАСАДИК Х.І.....	3
SHVEC V.V.....	5
NESTERENKO S., STANOVSKA T., STANOVSKYI A., DADERKO A.	6
PROKOPOVICH I., DUKHANINA M., DOBROVOLSKAYA V., NADERI BAKHER.....	7
STANOVSKYI O., SHVETS P., BONDARENKO V., TOROPENKO O., ABU SHENA OSA- MA, WALID HUSSAIN	8
FOSHCH T.	10
KIVA I.L., ALEKSASHIN A.V.	12
SAVELYEVA O., BEREZOVSKAYA E., MONOVA D., HEBLOV I.	14
MATICO F. D., ROMAN V. I., FEDORYSHYN R. M., PISTUN Y. P.	16
KOROTYNSKYI A., ZHUCHENKO O.	18
FEDORYSHYN R., SAVYTSKYI V., KLYOS S.	20
PALAGIN O.V., QASEM A.M., TKACHENKO O.M., KASIM M.M.....	22
BORIS V.V., ZHYGAILO O.M.	25
HEREGA A., KPYVCHENKO Y.	26
KHOBIN V.A., LEVINSKYI M.V.	27
DASCHENKO O., STANOVSKYI O., KHOMIAK Y., NAUMENKO E.	29
PELYKH S.N., ODREHOVSKA E.A., MAKSYMOVA O.B.	31
SIUMACHENKO D. M.	32
ЧАЙКОВСЬКА Є.Є., МАТВІЧКО Н.О.....	33
SHEIKUS A.R., LEVCHUK I.L., TRISHKIN V.YA., GYRENKO A.O.	35
TITLOVA O.	37
YEHOROV V., HABUIEV K., KUCHERENKO N.	39
YEHOROV V., PUTNIKOV D.	41
YEHOROV V., IHNATIEV S., SLUTSKYI D.	42
DROBIK E., TITLOVA O.	44
БЕЛОУС Н.В., КРАСОВ А. И.	46
VOLKOV V., POPOV A.	48
LOBODA Y. G.	50
ORLOVA O.U.	52
YEHOROV V., GOLUBKOV P.S., SHIPKO A.I.	53
ДМИТРИЕНКО В.Д., ЛЕОНОВ С.Ю.	55

ANALYSIS OF THE CLUSTERING OF RESULTS USING THE METHOD OF K-MEANS
AND WAYS TO IMPROVE THEM

For development of control algorithms of difficult objects methods of data mining (Data Mining), in particular – clusterization can find the effective application. It is known that methods of the cluster analysis are realized in rather large number of program tools from different producers. But estimates of results of their use continue to be considered not as absolutely satisfactory. Therefore development of advanced algorithms of the cluster analysis of data continues to remain an urgent task.

Before carrying out the cluster analysis three examples of a data set have been created:

- 1) «The 2-dimensional test» – 34 objects with two signs on values of coordinates of X and Y (4 clusters are provided);
- 2) «The 3-dimensional test» – 204 objects with three signs on values of coordinates X, Y and Z (8 clusters are provided);
- 3) «Flour - figures of merit» – 730 objects with three signs on values of number of falling, IDK, quantity of a crude gluten for the batches of flour arriving on bakery within ten months (the cluster design is unknown).

By the first two examples existence of such objects was in advance provided («disputable») which can belong to different clusters.

As a basic algorithm of a clustering the k-means method which is well described is chosen, it is applied in the majority of the known program tools and yields positive results. And Deductor Studio Academic, Statistica StatSoft, SPSS Modeler IBM became tools which were used for a clustering of the presented examples and carrying out the comparative analysis.

It is known that the choice of initial centers of clusters (centroids) has very significant effect on correctness of results of a clustering. As for the k-means method this procedure has no rigid determination and initial centroids are selected in a random way or on some special algorithm, it affected the accuracy of the received results on the first two examples. Any of tools didn't reach ideal filling of clusters, i.e. «disputable» objects appeared in «others» clusters.

The main criterion that determines the accuracy of hitting the object in the corresponding cluster, we used the value of the total distance between all objects and their centroids. This criterion can be considered sufficient to justify the assessment of the quality of clustering within a single analysis of the distribution of objects on a fixed number of clusters. Additional criteria appeared: Dan's index and index of assessment of a silhouette.

Therefore the software module realizing the procedure of an advanced clustering of data by k-means method was developed. Algorithms of automatic calculation of quantity of clusters and determination of provisions of initial centers of clusters (centroids) became its distinctive feature. At the same time, creation of polygons of frequencies on each sign was used.

The results provided the advantage of this module for all criteria.