

ABSTRACT

The thesis is presented in 51 pages. It contains 2 appendixes and bibliography of 23 references. Seven figures and 11 tables are given in the thesis.

The goal of this thesis is to choose mathematical methods and create software for information searching among data and data's attribute in initial multidimensional array data's characteristics in question.

In the thesis existing solutions of the problem are analyzed - methods of clustering - hierarchical, k-means, c-means, expectation-maximization (EM) algorithm, objective computer clustering (OCC) algorithm - for data, and regression analysis - linear, polynomial – for data's attributes. They are compared in terms of the complexity of algorithms, accuracy, efficiency and adequacy of methods when applied to multi-dimensional data. In the thesis, methods of clustering, such as hierarchical, k-means, c-means, are selected for data and multivariate polynomial regression analysis is selected for data's attributes.

For each input data set clustered original data set and regression function are formed. The automated system implementing the chosen method is developed. The developed system is tested.

Keywords: multidimensional data array, clustering, regression, information, cross-validation, correlation.