THE INFORMATIVELY-EXTREME DATA CLUSTERING OF CONTROLLED TECHNOLOGICAL PROCESSES

Radchenko T. A., student, group IN-53, Sumy State University
Supervisor: Shelehov I. V., assistant of informatics department, Sumy State University

The Informatively-Extreme Intellectual Technology (IEIT) further development prospect of the learning Control Systems (CS) synthesis consists in the development and inculcation of the hybrid algorithms of optimization of parameters of functioning of such systems. The author examines one of the possible variants of Data Clustering methods hybridization, which connects genetic algorithms and informatively-extreme methods within the framework of IEIT.

Let's consider Data Clustering (Cluster Analysis) as the classification of a great number of objects into different groups (classes), or more precisely, the partitioning of a data set into subsets (clusters), so that the data in each subset (ideally) share some common trait - often proximity according to some defined distance measure on the basis of a certain mathematical criterion of a partition quality.

Data Clustering contains such stages:

- selection of features (descriptions);
- determination of a distance measure;
- partition of a data set into subsets;
- search for the optimum partition.

The distance measure selection (measures of the object closeness) is considered to be a key moment in Data Clustering. The result of the object partition depends mainly on it. It determines the calculation of the similarity of two elements. In every concrete task this selection is made according to the importance (main goals) of the research, the physical and statistical nature of the used information etc.

The fitness-function in genetic algorithm is used as the quality criterion. It drives the calculation towards the best solution. It simultaneously is an inconvenience, which is related to the necessity of independent determination of a partition quality criterion, and also an advantage of the genetic methods, as it (such possibility) predetermines there high flexibility and there independence on structure and values of an entrance data.

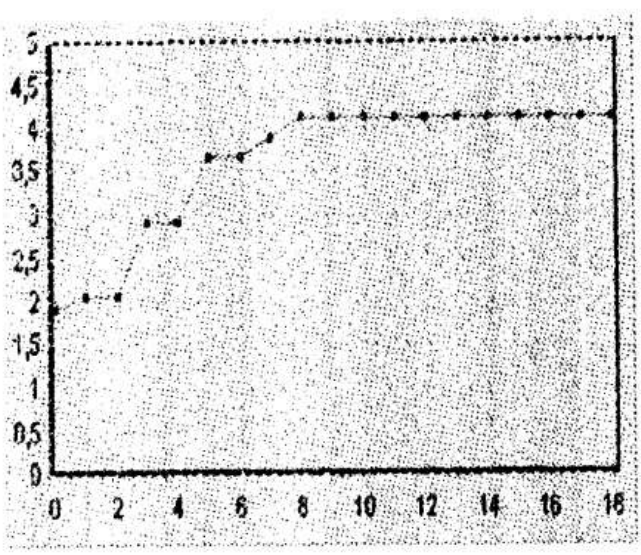
The Functional Efficiency Criterion (FEC), which is used in IEIT, is based on a direct estimation of the informative possibility of the CS and allows to set the optimum of spatio-temporal features of its functioning with the purpose to construct a faultless classifier in the discrete space of signs of recognition. The FEC informative nature allows it to be used as a universal distance measure in Data Clustering, as a fitness-function in genetic algorithms and as a measure of effi-

ciency of the CS functioning in general. The proper changes are also inserted into the categorical model of the learning process according to IEIT, especially to the basic contours of optimization, which necessarily include a term-set of values of the proper parameters and a term-set of the FEC values. A new contour is added to optimization and it is responsible for the CS functioning in the mode of Data Clustering and in the mode of genetic algorithm parameters selection.

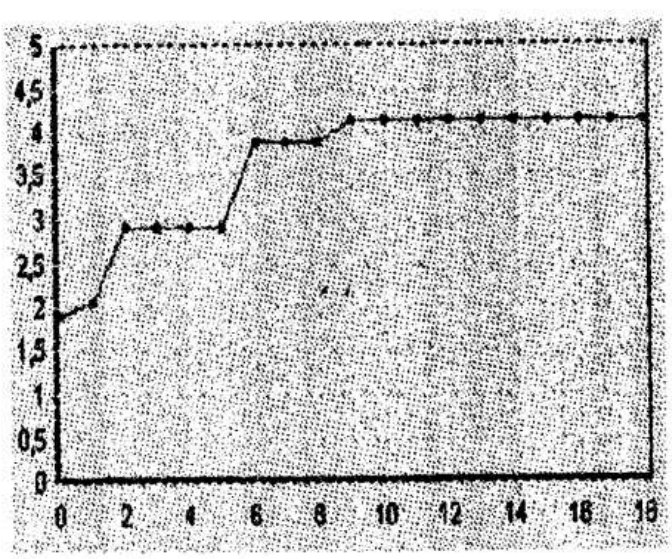
Also according to the peculiarity of a solving task (because of the combining difficulties of Data Clustering and IEIT and, as a result, the receiving of a desirable optimum of such combination) the algorithms of crossover and mutation were modified.

With the purpose of increase of productivity and speed of calculation of genetic algorithm it was decided to inculcate the mechanism of parallel calculations (the conception of islands). In accordance to this conception the population is divided into a few (in our case two) different subpopulations (demoses), which in future develop independently. Then on some stage on the basis of a random choice there is an exchange by some set of individuals between the subpopulations. And so it can proceed before the algorithm is completed.

The genetic algorithm calculation results are shown below on pic. 1-2.



Pic. 1 – The growth dynamic of FEC in population „Island I"



Pic. 2 – The growth dynamic of FEC in population „Island II"

Here population „Island I" and „Island II" obtained maximum of FEC valued 4,1658 on 8 and on 9 epochs appropriately. It proves efficiency of developed modifications of base genetic algorithm, quality of clustering and accuracy of formed classifier.

The algorithm was constructed on the base of the modified categorical model and successfully tested on the base of solving of practical tasks of the increase of technological processes control efficiency in public corporation „Sumykhimprom".

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ С ПОМОЩЬЮ MS OFFICE EXCEL 2007

Кунцев С.В., УАБД,г. Сумы

В современном бизнес-анализе получила распространение новая технология интеллектуального анализа данных - Data Mining. Она представляет собой технологию поддержки процесса принятия решений, основанную на выявлении скрытых закономерностей и систематических взаимосвязей между переменными внутри больших массивов информации [1]. Накопленные сведения автоматически обобщаются. Полученную информацию можно характеризовать как знания и применять к другим массивам данных. Обнаружение новых знаний используется для повышения эффективности бизнеса.

Методами Data Mining можно выявить определенные типы закономерностей, среди которых обычно выделяют пять стандартных типов: "ассоциацию", "последовательность", "классификацию", "кластеризацию", "временные закономерности".

С помощью методов Data Mining можно успешно решать сложные задачи в экономике. В банковском деле с их помощью выявлять мошенничество с кредитными карточками, разбивать клиентов на категории, прогнозировать клиентуру. В страховании – анализировать риски. В телекоммуникационном бизнесе – проверять лояльность клиентов.

Табличный процессор MS Office Excel 2007 обладает набором активных надстроек, которые позволяют решить сложные математические задачи. Возможности процессора можно существенно увеличить, если использовать надстройки для "интеллектуального анализа данных" [2]. В этих надстройках используются аналитические средства SQL Server 2005 Analysis Services [3].

Надстройки интеллектуального анализа данных для Office 2007 помогают выявлять закономерности и тренды, существующие в сложных данных, отображать такие закономерности в диаграммах и интерактивных средствах просмотра, а затем формировать цветные сводные отчеты для презентаций и бизнес-аналитики. Можно анализировать корреляции и формировать прогнозы для данных, хранящихся в таблицах Microsoft Office Excel, или создавать и изменять модели интеллектуального анализа данных, хранящихся в экземпляре Analysis Services, а также получать результаты в графическом виде в среде Microsoft Office Visio.

Надстройки интеллектуального анализа данных для Office предоставляют мастера, которые максимально упрощают получение полезных сведений из кубов или работу с большими базами данных.