Application of the clustering algorithm in an automated training system

A A Popov, A K Ovsyankin, M R Emomaliev and M M Satsuk
Reshetnev Siberian State University of Science and Technology, Krasnoyarsk, prospect "Krasnoyarsky rabochy" 31, Russia

E-mail: tolynbms@yandex.ru

Abstract. The development of information technologies provides a new, unique opportunity to conduct classes with the introduction of automated training systems for disciplines, which allows the student to choose the time and place for training, makes it possible to use new information technologies in training, and to a certain extent reduces training costs. The problem of improving the quality of the educational process in an educational institution is being solved because of the use of automated systems. This task is the main one in the activities of each educational institution. The development of a high-quality system will provide an increase in the student's knowledge in the chosen field, and help to form an up-to-date curriculum for training. The purpose of this work is to create an automated training system using the clustering and content ranking algorithm, which can improve the quality of the student's educational process.

1. Introduction
The constant increase for information and the limited study time necessitate the intensification of training, the development and implementation of non-traditional technologies based on the use of computer technology using active teaching methods in all their diversity and complexity. The implementation of active teaching methods is one of the main tasks of didactics, which involves the activation of the entire process, the identification of a system, methods, techniques that contribute to an increase in the activity of students through the formation of a positive motivational structure of educational and cognitive activity [1].

The development of information technology has provided a new, unique opportunity to conduct classes with the introduction of automated training systems for disciplines in universities. It, firstly, allows the trainee to choose both the time and place for training, secondly, it makes it possible to use new information technologies in training, and fourthly, to a certain extent, it reduces training costs. On the other hand, the introduction of new automated teaching systems into education enhances the possibilities of individualization of teaching [2].

Now, many different training systems solve the problem. One of such systems is the "Base and Generator of Educational Resources" hereinafter referred to as BiGOR.

BiGOR implements a method of creating educational materials based on "assembly" from pre-developed modules and shared content units [3]. Shared content unit technology is a principle that ensures the relationship between elements by affixing relationships.

However, most of them are narrowly focused systems, which include the methodological and curriculum of the learning process only in a certain area. The proposed approach to the use of clustering
and ranking algorithms will create a training system that can improve the quality of training and select a training program based on the interests of the user and his curriculum.

With the development of distance learning technologies, the development of such systems today is relevant and in demand.

2. Methods
Automated training systems (AOS) are designed to accompany the educational process through software and technical support. Such systems include methodological, educational and organizational support of the process [4]. Figure 1 shows the classification of AOC [5].

Now, the following functions of AOS can be distinguished, which are useful in the educational activities of students:

- Development of training courses.
- Determination of the initial level of knowledge and progress of the student.
- Performance monitoring.
- Statistical analysis of the learned material for each course and trainee.

![Figure 1. AOS classification.](image)

The developed system offers the user an AOS of a combined type, which contains elements of an information, reference and training system using a clustering algorithm and content ranking according to the user's interests. Figure 2 shows the structure of the AOC.
In the system being developed, it is planned to cluster the database of educational materials. Let us introduce the definitions of the concepts with which we will operate.

Object is an elementary data group with which clustering algorithms operate. Each object is identified with a vector of characteristics: \( x = (x_1, ..., x_d) \). The \( x_i \) components are separate characteristics of the object.

A cluster is a subset of “close to each other” objects from a set. The distance \( d(x_i, x_j) \) between objects \( x_i \) and \( x_j \) is the result of applying the chosen metric in the space of characteristics [6] [7].

Clustering is the principle of dividing a set of objects into groups, called clusters. Each of the created groups should contain similar objects, and the groups should be as different as possible from each other. Thus, the clustering of the training system data can be divided into the following stages:

- Retrieving objects for clustering.
- Obtaining criteria for evaluating objects.
- Calculation of the measure of similarity between elements.
- Using the principle of cluster analysis.

When dividing multiple objects into groups, you need to identify similar objects and group them. In this system, this can be achieved by measuring the distance measure. For each object, a vector of characteristics is generated with the help of which the "degree of similarity" of the data is subsequently calculated using the metric in the Euclidean space.

Euclidean metric is the distance between two points of Euclidean space, which is calculated by the Pythagorean theorem:

\[
\rho(x, x') = \sqrt{\sum_i^n (x_i - x'_i)^2}
\]  

The result of the work of the hierarchical clustering algorithm is a dendrogram that allows you to split the original set of objects into any number of clusters. This clustering algorithm builds a partition of clusters “bottom-up”, that is, at each step; it combines two clusters with the smallest distance between any two representatives (figure 3).

The advantage of hierarchical clustering is that you can try to determine the required number of clusters by examining the properties of the resulting tree, for example, to select into different groups those subtrees, the distances between which are large enough. It is convenient to work with the resulting structure to find clusters in it. Conveniently, such a structure is built once and does not need to be rebuilt when searching for the required number of clusters.
3. Results
Based on the developed UML use-case diagram (figure 4), the ER-diagram of the database is capable of storing the entire amount of data of the future AOC. Includes the following tables (figure 5).

Figure 3. An example of the Single-link algorithm.

Figure 4. UML use case diagram.
In order for the developed IS to be user-friendly and to have the most complete functionality, it is necessary to adhere to the following requirements:

- Simple and intuitive interface to increase the speed of the user’s work by reducing the time of thinking;
- Reducing the number of human errors by reducing the requirements for vigilance, increasing the legibility and visibility of indicators, blocking potentially dangerous user actions until confirmation of the correct action is obtained;
- The user interface should contain tips, informational messages, and help documentation.

Based on the above criteria, the interface of the future information system was developed (figure 6). The color scheme consists of various shades of gray, white and black, which does not create unpleasant eyestrain. The user should not feel tension while working in this system [8].

The main menu of the training system page consists of functional blocks with the main categories of academic disciplines that users use most often.
4. Conclusion
In the course of the work performed, an automated training system of a combined type was designed, which gives the student access to the database of training materials, and also, using the data clustering and ranking algorithm, offers the user filtered information based on his interests and preferences. In addition, a layout of the interface of the main window of the automated training system has been developed.

In the future, it is planned to create an automated training system using the clustering and content ranking algorithm.

References
[1] Pak N I 2004 Nonlinear teaching technologies in the context of informatization: Textbook (Krasnoyarsk: RIO KGPU)
[2] Mogilev A V 1999 Informatics (Moscow: Academy)
[3] Karpenko A P and Sokolov N K 2008 Extended semantic network of the learning system and assessment of its complexity Science and Education: scientific publication of the Moscow State Technical University. N.E. Bauman B11 4
[4] Bolshakov A 2006 Intelligent automated training systems: methodology for creating Higher education in Russia 6
[5] The purpose and functions of AOS, the tasks they solve, AOS classification https://studwood.ru/1676762/informatika/naznachenie_funktsii_reshaemye_zadachi_klassifikatsiya
[6] Kotov A and Krasilnikov N 2006 Data clustering New information technologies in automated systems 17
[7] Knyaz Dimitri 2014 Analysis of the main algorithms for clustering multidimensional data (LAP)
[8] Cooper A and Reiman R 2009 Interaction Design Basics (Symbol-Plus)