Mathematical model concept for education progress control system

I.F. Astachova[a], E.I. Kiseleva[b]

[a] *Voronezh State University, 394006, University Square 1, Voronezh, Russia*
[b] *Voronezh State Pedagogical University, 394000, Lenin Street 84, Voronezh, Russia*

Abstract
The scope of the article is dedicated to the design of the oretical and hands-on investigation of mathematical model concept for Education Progress Control System (EPCS) that might be applied for educational institutions. The versatility of the EPCS lies in its independence from any specific subject content (or related factors), which allows teachers to upload the required course directly to the educational platform, and students to choose an individual courses in order to plan their educational programs for their own. To implement the designed EPCS the hybrid system design approach was chosen, because it combines the advantages of different technologies that allow solving complicated tasks. The well-known fuzzy logic design allows determining the educational content sections, but this approach causes significant difficulties for students. However, the implementation of Kohonen self-organized network (or Self-Organizing Map (SOM) is a computational method for the visualization and analysis of high-dimensional data, especially experimentally acquired information) allows classifying students involved into educational programs provided by the EPCS platform.

Keywords: mathematical model of learning and control system; information system; training course

1. Introduction

In recent years, various concepts of IT-based educational platforms have spread out. According to the data available in the literature, the main differences in most cases are following: the distribution degree of control functions between the system's user and the system itself (for some cases user is allowed to choose the scenario of his/her interaction with the system, for other cases the scenario is fully or partially implement into the system and cannot be changed by the user); the amount of integrated content – theoretical courses and self-check tasks (and their combination implemented into the system); the presence or absence of education progress control function [1,2]. The designers of such educational systems have chosen the various combination of parameters and the result "melting pot" has been the deliverable of the proposed educational platform. However, all almost all of the currently known EPCS platforms have the potential to accumulate statistical information regarding the students' educational scenarios (the courses they choose, the order that the courses were taken by them), typical mistakes that take place during students' self-checks, the overall amount of passed and failed courses, etc.

The main purpose of this work is to design the Education Progress Control System (EPCS) concept that might be applied for educational institutions and might be applied for students' education and knowledge check needs. The one of main objectives is to develop a mathematical model for such EPCS and to choose the oppropriate means for its optimization. In order to develop the EPCS we have chosen the hybrid system (HS) design approach, because it combines the advantages of different technologies that allow solving complicated tasks. HS is a system that combines two or more different computer technologies [3]. We have taken the parametric model of common educational process created by V.M. Monakhov [4] as a basis for the EPCS mathematical model. The parameters that holistically show the patterns of the educational process are following: goal setting parameter (the designed system of micro-goals); diagnostics parameter; educational dosing parameter for certain student's activity; the logical structure parameter of the educational project; correction parameter. The listed parameters are forming the basic design approach for the EPCS mathematical model.

The provided EPCS concept sets the purpose to split each particular educational section and its content on the micro-goals, and stores the fact of students' achievements, using the self-checks tasks embedded into the system. The self-checks itself is a set of mandatory training tasks that are required to be performed by a student. If a self-checks has not been performed, the correction is carried out, and then followed by the new self-checks test. The dosing of individual students activity is a set of content that is offered for students during the development of certain training section.

2. The system model

The Figure 1 shows the EPCS concept schema by the embedded components.

The mathematical model of the EPCS can be presented as following:

$$URS = \langle S, PP, T, CK, RR \rangle \quad (1)$$

where S – is the educational process; PP – the profiles factor of teachers and students; T – the educational path; CK – relates to the model of the current knowledge of certain student; RR – the profile of the educational course [5]. To store the personal information of teachers and students the profile entities are used. The educational path is represented by the information about the students' the educational path scenario within the EPCS.

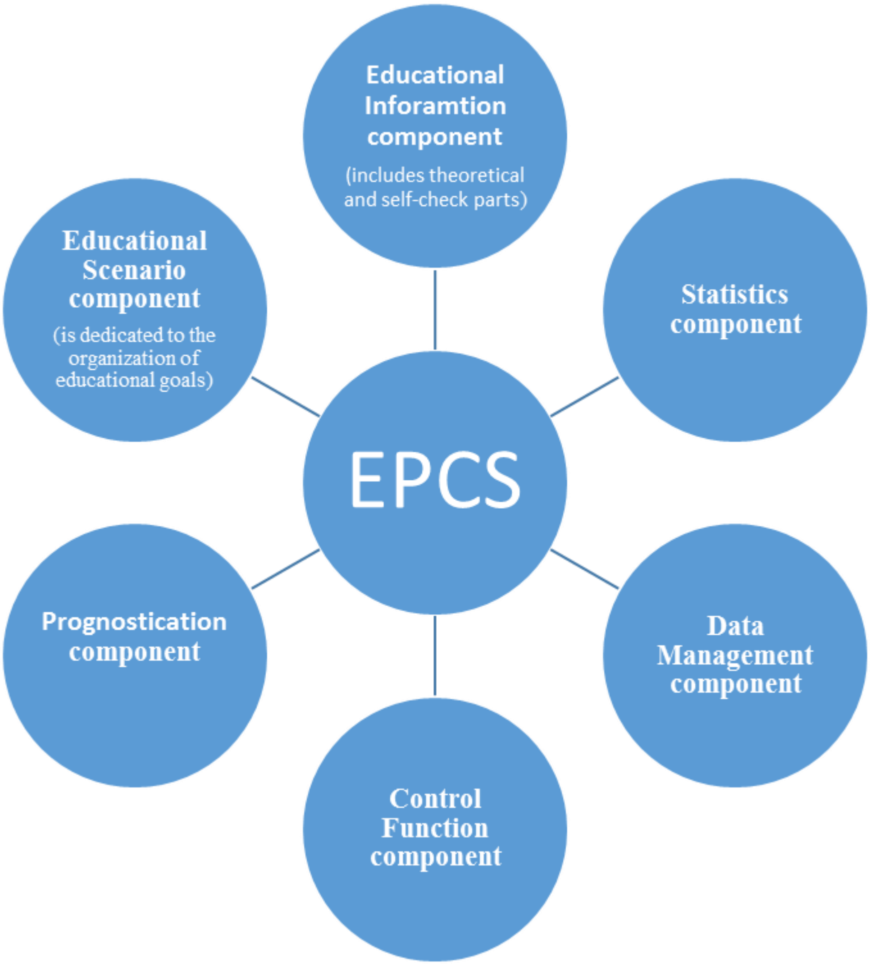

Fig.1. The model of hybrid learning system.

In accordance to the parametric model of the educational process proposed by V.M. Monakhov [4], we have developed the following model of the educational process in the EPCS according hybrid system design approach (S)

$$S = \langle G, CT, EM, Q \rangle \quad (2)$$

where *G* – the goal setting parameter, represented by the EPCS's educational micro-goals; *CT* – the educational courses branch, that corresponds to the logical structure of educational process; *EM* – the educational content, used for the correction and educational dosing of for the independent students; *Q* – the educational diagnostics parameter, comprising a plurality of self-check tasks. The set of educational objectives can be presented as following mathematical model:

$$G = \langle GN, GS \rangle \quad (3)$$

where G – the educational branch of certain educational micro-goals; GN – the set of educational edges for branch's dedicated educational objectives; GS – the set of branch's edges, reflecting the hierarchical relationship between certain educational objectives. The learning content can be determined by using the equation below:

$$EM = \langle P, E \rangle \quad (4)$$

were *P* – the set of educational pages; *E* – the set of educational elements.

The self-check or knowledge diagnostics parameter Q might be defined as a set of self-check tasks. The set of the self-check questions (*Q*) is defined by:

$$Q = QS \cup QB, \quad (5)$$

where *QS* – the multiple sets of self-check questions; *QB* – the set of test questions sections.

To describe the content of certain educational course content we should use the following equation to determine the educational branch:

$$CT = \langle CTN, CTS \rangle \quad (6)$$

where *CT* – the branch of the educational course content; *CTN* – the set of branch vertices, points of content; *CTS* – the set of branch edges, reflecting the relations between the involved educational content elements.

The components are models ***G***, ***CT***, presented in the form of graphs, which can be optimized by using the appropriate well-known algorithms, for example:

CT – the branch of the educational material; *CTN* is the set of branches vertices, the educational content items; *CTS* – the set of the educational branch edges, reflecting the relations between the involved educational content elements. The content of the educational material represents as orientated graph [8, 9]. As the statistics weights of the edges the coefficient a_{ij} of difficulty transition from the vertexes ***i*** to ***j*** of the educational branch might be applied. Initially, these coefficients can be empirically determined by the peer review and then adjusted by using of well-known math-numerical methods. Thus, the problem of determination of the optimal educational sequence for certain educational content can be reduced to the task of optimizing of the related orientated graph [10].

3. Methods

The EPCS concept is divided on the components that are split on several sections (please refer to the Fig. 1 above). The main EPCS component is the algorithm that builds the educational branch (**Educational Scenario component**), and the components of this branch can be divided on micro-goals. The **Prognostication component** with the use of fuzzy logic allows determining the educational sections, which cause some difficulties for students (bad self-check statistics, long investigation periods for theoretical parts). Therefore, the fuzzy logic approach allows adjusting the micro-goals content as well. The **Educational Information component** includes theoretical and self-check materials and allows determining the dosing parameter during certain educational scenarios. The **Control Function component** directly corresponds with the diagnostics parameter of educational process.

The implementation of Kohonen self-organized network allows classifying students involved into educational programs provided by the EPCS platform by three parameters: students that have not passed the self-checks and have not finished the educational scenario; students that demonstrating the knowledge level, corresponding to the current educational standards; ; students that demonstrating the knowledge level above the current educational standards. The **Statistics component** allows storing information about the educational progress for students who are registered in EPCS platform and these data is provided by Data Management component.

To classify students we use Kohonen self-organized network technology. The classification itself is the decomposition of objects to several sets, the number of which can be known in advance. This network consists of a single layer of neurons and learns independently (based on self-organized neuron map). In this case, an existence of a certain number of data classes is expected. The class can be prototyped as the vector data that is mostly typical for each class. Therefore, for each selected data vector the closest prototype can selected (and this process can be iterated across other data vectors), then the centroid composition of all vectors (connected with the original prototype [2]) becomes a new prototype for each class.

The Kohonen self-organized network is represented as number of inputs for each neuron that is equal to the dimensional parameters of the classified object. In our case, the results of work on certain educational scenario of certain students will be classified. As it was determined above in the article, each individual work on the educational scenario consists of four tasks – therefore, each neuron has four inputs correspondently. The number of neurons equals the number of groups that students are divided. In our case, we can divide students into four groups as following:

1. Have not passed the self-check (these students will be directed to the correction group);
2. Have shown knowledge level that corresponds to the state educational standard requirements (these students will be directed to the satisfactory group);
3. Have shown knowledge level that is above or equal the state educational standard requirements (these students can be assessed as "good");
4. Have shown the outstanding knowledge level that is above the educational standard requirements (these students can be assessed as "excellent).

However, in case of necessity the number of listed above classes can be changed by changing of the neurons number for certain neuron network map. During the initial network training, the initial values of weight matrix were set randomly so that are small values. Upon the presentation of the training sample, weight matrix is modified durng self-organization process. Each column of the weight matrix represents itself as set of parameters for corresponding neuron-classifier. For each j-neuron ($j = 1, 2, .., m$) the distance might be determined from the single neuron to the input vector X:

$$d_j = \sum_{i=1}^{n} (x_j - w_{ij})^2. \qquad (7)$$

Then the neuron with the number $k, 1 \le k \le m$ is chosen – for this neuron the distance is minimal, (i.e., the network classifies the input vector to class number *k*). Therefore, at the *N*-step of training only the weights for neurons in the vicinity of neuron k will be modified:

$$w_{ij}^{N+1} = w_{ij}^{N} + \alpha_N \left(x_i - w_{ij}^{N} \right). \qquad (8)$$

where α_N – the abstract pace of learning, expressed in number.

Initially, in the vicinity of any of neuronal neuron network can be located, but with each step further this vicinity is shrinking. Thus, at the end of the training iteration only the weights of neuron with number *k* are participating in the learning process. After some time period the learning pace α_N also decreases (often considered [8] α_N =0.9, $\alpha_{N+1} = \alpha_N - 0,001$).

Images of the training sample are presented sequentially, and each time there is adjustment of weights.

The Kohonen network learning algorithm consists of the steps listed below:

1. Network initialization;
2. Assignment the network coefficients $w_{ij}, i = 1, 2...n; j = 1, 2...m$ as small random values to all weight involved. Values preset: α_N – initial learning phase and D_0 – maximal distance between the weight vectors (column of matrix *W*);
3. Submission of the new input signal X to network;
4. Calculation of the distance from input *X* to all neurons of the network:

$$d_j = \sum_{i=1}^{n} (x_i - w_{ij})^2, \; j = 1, 2...m \qquad (9)$$

5. Selection of neuron $k, 1 \le k \le m$ with the minimal distance d_k from the input point that is applied to all neurons of the network.

6. Adjustment of weights of k -neuron and all neurons that are located within a distance that not exceeds D_N :

$$w_{ij}^{N+1} = w_{ij}^{N} + \alpha_N (x_i - w_{ij}^{N})^2. \quad (10)$$

7. Decrease of values α_N, D_N .

Steps 2-6 might be iterated until the weights will stop shifting their values (or until the total values shift for all weights will be very small).

After the network has been trained, the classification of students' knowledge is carried out by applying the test vector to the network input and calculation of the distance from this input pint to each neuron (with further selection of the target neuron) with the smallest distance (as an indicator of the correct classification). This classification approach allows data fulfilling for the network, and can be based on the educational content of the verification work.

To increase the effectiveness of educational scenarios, diagnostics, and correction according to the objective in order to determine the educational course sections that are representing the difficulties for students the methods of fuzzy logic were used.

The educational prognostication is a quantitative estimation of the future condition of an object or system that is based on scientific methods. To predict the difficulties that students might face during the pass of educational scenario the methods of fuzzy logic were used that allow finding a fuzzy solution. That approach meets the requirements of the level of accuracy.

The statistical data results obtained from 2011 to 2015, during testing the 1-st year students of Psychological - Pedagogical Faculty (PPF) in the Voronezh State Pedagogical University are summarized in general table (please refer to **Table 1**). All data is presented in the general way (in order the one can compare and use it easily). During the testing period (2011 – 2015) various number of students were tested, the percentage of correct answers (during self-checks) for each topic is under review.

Table 1. The percentage of the correct answers on sel-checks.

№	Topic	2011	2012	2013	2014	2015
1	Definition and methods of set setting	10,53	30,00	38,10	43,18	54,35
2	Set manipulation	34,21	40	52,38	54,55	54,35
3	Parameters of set manipulation	44,74	55	66,67	70,45	73,91
4	Adaptability	86,84	65,00	59,52	77,27	69,57
5	Binary relations	55,26	42,50	73,81	75,00	76,09
6	The relations of equivalence and order	31,58	42,50	59,52	65,91	65,22
7	Combinatorics elements	89,47	80,00	85,71	93,18	86,96
8	Classical definition of probability	94,74	85,00	92,86	84,09	76,09
9	The probability of events sum and difference	57,89	57,89	83,33	93,18	91,30
10	Classical probability theory	36,84	40,00	61,90	86,36	91,30

The analysis of the data above has shown that the information provided does not allow identifying the educational course topics that requires attention. The solution of this problem is possible by means of the theory of fuzzy sets and fuzzy logic methods [3, 6].

The used prognostication algorithm consists of the following steps [7, 8, 9] listed below:

1. The variations calculation of the self-check correct answers percentage for each topic of the educational course as the difference between the self-check correct answers percentage during the current and previous years. Definition of the universal set U, which is an interval between the lowest and the highest percentage of correct answers variations on educational course theme: $U = [V_{\min} - D, V_{\max} + D]$, where $V_{\min}$ – is the lowest percentage of correct; $V_{\max}$ – the highest variation.

2. The division of the universal set U to several equal-length intervals, including various values of variations $U = \{u_i\}$ and definition of the intervals middles u_i^{mid} .

3. The introduction of a linguistic variable and definition of the corresponding linguistic values $A = A_i, i = 1, 2, ...m$,i.e. definition a plurality of fuzzy sets *F(t).*

4. The fuzzification of input data, i.e. quantitative transformation of constant values into the fuzzy values. This operation allows reflecting the appropriate qualitative conceptions, corresponding to the quantitative values, about the right answers in the group under the review values understanding of the responses in this group.

5. Parameter selection1<w<L, corresponding to the length of time preceding the current year, where L – the total number of years considered in calculations.

6. The calculation of fuzzy relations matrix:

$$R(t) = O^w(t) \cap K(t) \qquad (11)$$

where $O^w(t)$ – operational matrix, comprising of A_t as lines $t \in [k-w, k], K(t)$ – matrix criteria of the dimension $1 \times m$ for the forecasting year *t*.

7. Defazzification of the obtained results, i.e. the transition from the fuzzy values to the constant values (quantitative).

As a linguistic variable, the variable "variation of the percentage of correct answers on the educational topic" was used. It has the following values: {(negative significant variation), (negative small variation) (without changes), (a small positive variation), (a significant positive variation)}. The obtained results are used to optimize the educational scenarios and students' educational objectives of the EPCS platform, an increase the number of educational content of the course sections, which, according to the educational prognostication component, are produce significant difficulties for students. The informative part of the course is not taken into account for prognostication component, because the resulting algorithm can be used for any educational course, which can be included into the EPCS platform.

4. Results and Discussion

The following user interface (UI) has been designed for the implementation of the EPCS platform.

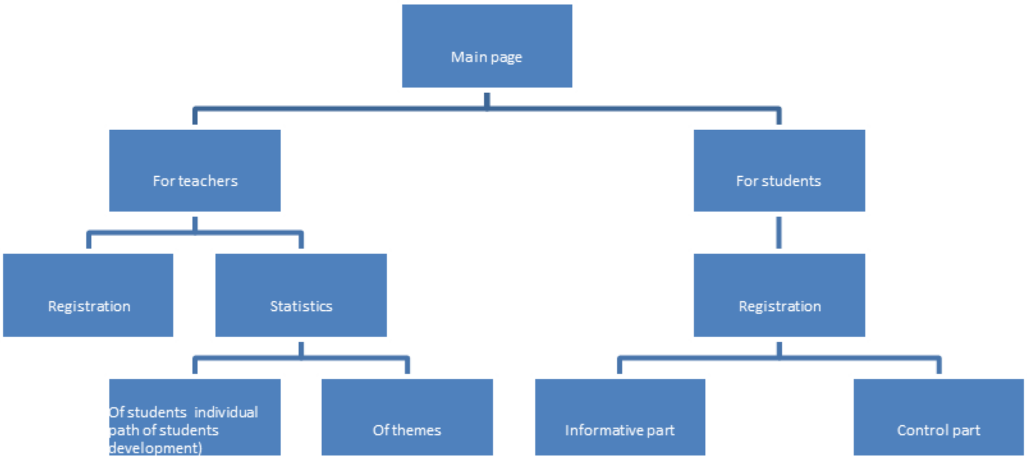

Fig. 2. The EPCS platform user interface.

The main page is common location on the UI for all users and represents itself as a connecting node between various sections of the System. The page provides two entry blocks - login into the teacher's account and login as a student. The users also have the opportunity to register a new account as a student or teacher [6, 7].

Once registered under the student account the user is immediately transferred to the **private student cabinet.** This page is the next binding node - being in the account, one can explore the theoretical section of the educational course material; perform practical work, pass the sel-cheks. If the student has been classified by the system as one who does not show the level of knowledge required by the state educational standards, he will be asked to go through the knowledge correction procedure, and then will be re-assessed. Students can view all of their results with a full provision of information about the tests passed, personal data change, etc.

After registering as a teacher, the user enters the personal teacher's cabinet. The user is immediately invited to page "Students", which is designed as summary data table of all students and their tests passed.

The EPCS platform described in this article, can be used during the process of distance learning for the organization and control of individual work of students, allowing to conduct efficient training and students' knowledge control on any educational course.

5. Conclusion

The mathematical model of hybrid learning and control system, considered in the article, provides the opportunity to optimize the learning process on various parameters using statistical information obtained in the course of system operation, which contributes to the structure of the system parameters, presented in the form of various graph.

References

[1] Jurkov, N.K. Intellektualnie kompyuternie obuchajushie sistemy. – Penza: Iz-vo PGU, 2010. – 455p.

[2] Clark, S *Assistive Software for Disabled Learners*./ S. Clark ., J. Baggaley . - Athabasca, 2004 – 320p.

[3] Batyrshin, I.Z. Nechetkie gibridnye sistemy. Teoriya i praktika / pod red. N.G. Yarushkinoy. – Moskva: FIZMATLIT. – 2007. – 208 p.

[4] Monakhov, V.M. Tekhnologicheskie osnovy proektirovaniya i konstruirovaniya uchebnogo protsessa. – Volgograd: Peremena, 1995 – 540p.

[5] Gapanyuk, Yu.E. Issledovanie i razrabotka modeli, metodiki i sredstv sozdaniya avtomatizirovannykh uchebnykh posobiy s ispol'zovaniem tekhnologii XML [Elektronnyy resurs]: Dis. ... kand. tekhn. nauk : 05.13.17. , Moskva, 2006

[6] P.Fritze, A. Ip Learning Engines a Functional Object Model for Developing Learning Resources for the www // ed-media & ed-telecom 98. -Freiburg, (1998), 53-58.

[7] Mel'nikova, A.A. Instrumental'nye sredstva modelirovaniya uchebnykh mul'timedia kompleksov - diss. na soiskanie uchenoy stepeni kandidata tekhnicheskikh nauk, - Samara, 2004.

[8] Roberts, F.S. Discrete Mathematical Models, with applications to social, biological and environmental problems. Prentice-Hall, 1976

[9] Holland, J.H. Adaptation in Natural and Artificial Systems, - Ann Arbor: The University of Michigan Press, 1975 – 120p.

[10] Ahuja, R. K. Applications of network optimization./ Ahuja, R. K., T. L. Magnanti, J. B. Orlin, and M. R. Reddy. - Cambridge, MA: Sloan School of Management, - 120p. .