Informatics and Control Problems 44 Issue 2 (2024)
journal homepage: www.icp.az

Automatic generation of concept maps for teaching informatics in higher education institutions

A.E. Aliyeva

Institute of Control Systems, Baku, Azerbaijan

A R T I C L E I N F O

Article history:
Received 14.11.2024
Received in revised form 21.11.2024
Accepted 06.12.2024
Available online 25.12.2024

Keywords:
Subject of informatics
Adaptive learning systems
Apriori algorithm
Concept map

A B S T R A C T

In the article, a method for automatic building of concept maps based on data mining technology for adaptive learning systems in teaching informatics in universities is applied. The proposed method is important for more effective presentation of educational material in adaptive learning systems.

1. Introduction

As technology develops rapidly and becomes an integral part of our lives, informatics plays an important role both in everyday life and in the work environment. Since the process of informatization is taking place in all areas of human activity, it is important for everyone to have high information literacy. As every specialist receives the initial basic knowledge in university, improving the learning process is an urgent issue. Currently, the development of informatics is in the spotlight all over the world. Thus, in recent years, institutes and research centers have been established in many countries to study and conduct research on informatics problems in the world. One of the main issues facing them is to propose modern educational standards in teaching the subject [1].

The recent widespread use of distance learning in higher education requires developing adaptive learning systems. Adaptive learning systems are widely studied in distance learning and in the field of intellectual education [2-10]. In these systems, a concept map is used as a tool for effective visualization of knowledge. A concept map is a very good tool for structurally presenting the topics, subtopics of a subject and the connections between them. A concept map is widely used in the field of education. The main purpose of concept maps is to better understand knowledge and connect topics. This, in turn, allows for accurate identification of students' knowledge gaps.

The application of concept maps to improve the quality of teaching informatics in adaptive learning systems in universities is a topical issue, because concept maps allow students to identify topics that they do not master well and to target those topics.

Building concept maps to improve the quality of students' learning process is an important research topic in adaptive learning systems [11-13]. Shyi-Ming Chen used the Apriori algorithm (Agrawal & Srikant, 1994) to build automatic concept maps [14-15].

E-mail address: aynur.aliyeva8020@gmail.com (A.E. Aliyeva)

www.icp.az/2024/2-08.pdf https://doi.org/10.54381/icp.2024.2.08

Studies show that this visual structuring helps students to understand the connections between topics more deeply. Therefore, in the article, we apply a method for building an automatic concept map to improve the quality of teaching informatics. It should be noted that software has been developed for conducting experiments to build an automatic concept map in teaching informatics in universities.

2. Method for automatic building of a concept map in teaching informatics

In the article, we present the application of the method for building an automatic concept map in adaptive systems in teaching informatics. First, the weights of concepts (topics) in test questions are assigned by the teacher. Each student's answers to the questions are recorded. Then, based on the students' answers to the questions, we apply the Apriori algorithm based on four types of association rules. Finally, to build a concept map in adaptive systems, we calculate the degree of correspondence between any two concepts. The degree of correspondence tells the student which concept should be studied before a certain concept.

Let us show students' answers to certain test questions and the relationship of each question to certain concepts (topics) in the form of a matrix. Note that a question can be related to several concepts. Each question has a weight for the concepts it is related to, and the sum of these weights equals one.

Suppose that n students gave $T_1, T_2, \dots, T_m$ answers to $L_1, L_2, \dots, L_n$ m test questions. Then the scoring matrix G is given as follows:

$$G = \begin{bmatrix} g_{11} & \cdots & g_{1m} \\ \vdots & \ddots & \vdots \\ g_{m1} & \cdots & g_{mn} \end{bmatrix},$$

where $g_{ij} \epsilon [0,1]$ indicates the incorrect answer of student L_j to test question T_i as $g_{ij} = 0$, and the correct answer of student L_j to test question T_i as $g_{ij} = 1$, $i = 1,2,\dots,m,\ j = 1,2,\dots,n$. The weight matrix between test questions and concepts will be as follows:

$$TC = \begin{bmatrix} tc_{11} & \cdots & tc_{1n} \\ \vdots & \ddots & \vdots \\ tc_{m1} & \cdots & tc_{mp} \end{bmatrix},$$

Where tc_{ij} denotes the weight of the concept C_j to which the question T_i belongs, $0 \le tc_{ij} \le 1,\ i = 1,2,\dots,m,\ j = 1,2,\dots,p$. The proposed method for building a concept map based on the answers to the test questions consists of the following steps.

The algorithm that determines association rules between the elements. Based on the student's answers to the test questions and the Apriori algorithm (Agrawal and Srikanth, 1994), the following association rule is determined between the questions:

1. If a student answered question T_i correctly, that person also answered question T_j correctly.
2. If a student answered question T_i incorrectly, that person also answered question T_j incorrectly.

For each question pair, we create association rules based on students' correct and incorrect answers for each question in that pair and calculate their support value. Then, new question sets are created, each consisting of two questions. We denote these question sets by $(T_i \to T_j)$. For each question pair $sup(T_i \to T_j),\ \ i \neq j, i = 1,2,\dots,m,\ j = 1,2,\dots,m,\ I$ confidence value is calculated:

$$I(T_i \to T_j) = \frac{sup(T_i \to T_j)}{sup(T_i)}, \tag{1}$$

where, first, the support values $sup(T_i)$ are calculated for all questions T_i $i = 1,2,\dots,m,$, one by one, and the questions that satisfy the minimum support are selected. Assume that questions have a minimum support value of 40%, that is, 1-element answers must have a support value of at least two.

Finding relationships between concepts. We calculate the degree of relationship between concepts in question pairs based on the concept weight assigned by the teacher for each question. If there is more than one relationship between any two concepts, then we take the one with the highest relationship degree.

For questions with correct and incorrect answers, we calculate the relationship between concepts, using the following formula:

$$R(C_i \to C_j)_{T_xT_y} = \min\left(W_{T_xC_i}, W_{T_yC_j}\right) \times I(T_i \to T_j), \tag{2}$$

where $R(C_i \to C_j)_{Q_xQ_y} \in [0,1]$ denotes the degree of relationship between the concepts "$C_i \to C_j$" and the questions $T_x \to T_y$, $i = 1,2,\dots,p,$ $j = 1,2,\dots,p,$ $x = 1,2,\dots,m,$ $y = 1,2,\dots,m.$ Here, $W_{T_xC_i}$ denotes the weight of the concept C_i in the question T_x, $W_{T_yC_j}$ the weight of the concept C_j in the question T_y.

After finding connections, we combine the relationship degrees for correct and incorrect answers for concepts C_i and C_j in one table, satisfying the following conditions:

- If there is a relationship in the "$C_i \to C_j$" relationship table, we write it in the combined relationship table.
- If there are two relationships in the "$C_i \to C_j$" relationship table, we calculate the relationship difference with the following formula and write it in the combined relationship table:

$$R' = \frac{R(C_i \to C_j)^+ - R(C_i \to C_j)^-}{MAX(R(C_i \to C_j)^+,\ R(C_i \to C_j)^-}, \tag{3}$$

where $R(C_i \to C_j)^+$ is the degree of conceptual relationship of correct answers, $R(C_i \to C_j)^-$ $i = 1,2,\dots,p,$ $j = 1,2,\dots,p,$ is the degree of conceptual relationship of incorrect answers. The table is constructed within the following conditions that the quantity R' satisfies:

- If the difference is greater than a certain number $\mu, 0 \le \mu \le 1$ set by the user, we delete the relationship between the concepts C_i and C_j. Otherwise, we write the one with the largest degree of correspondence between the concepts C_i and C_j into the relationship table.
- If the confidence level between concepts C_i and C_j is smaller than the number set by the user, we delete it.

We build a concept map based on the confidence values we calculated according to the association rules and the table obtained within the conditions of the R' quantity listed above.

3. Automatic building and analysis of a concept map

3.1. The importance of the concept map in teaching informatics

Concept maps play an important role in teaching informatics. Concept maps are useful for facilitating the learning process for students and for assimilating information more efficiently. The main advantage of concept maps in teaching informatics is that they help to structure the complex content of the subject.

Note that computer systems, algorithms, and programming languages can be divided into smaller subtopics and displayed on a concept map. This allows students to easily grasp the main topic and relate each subtopic to the overall concept. It also allows students to identify concepts (topics) that they have not mastered and recommends the order in which they should study the topics. Understanding the relationships between the concepts (topics) of informatics can sometimes be challenging.

Let us look at the description of the method on some topics of informatics. Areas such as programming, databases and algorithms require an understanding of the working principles of computer systems. The concept (topic) of “Computer Systems” covers concepts such as the main parts of a computer, hardware and software, operating systems, file structures. Therefore, the student should not study the concept of “Programming” before the concept of “Operating Systems”. The concept (topic) of “Operating Systems” is related to computer systems and covers the operation of software, as well as system programs (drivers, security programs, etc.). The concept of operating systems also provides a link to the topics of data security and networks. The concepts of “Programming” and “Algorithms” include the concepts of programming languages (Python, C++, Java, etc.), basic programming concepts (variables, loops, functions), properties and types of algorithms. The concept of “Cybersecurity” includes the principles of information security, encryption, authentication, threat analysis and defense mechanisms. This concept is related to the topics of networks and operating systems, because in order to understand the risks and defense methods in this area, it is necessary to understand the working principles of computer systems. Cybersecurity also requires learning the concepts of coding for security in the field of programming. As can be seen, the concept of “Operating Systems” should be studied first, then “Algorithm”, “Programming”, “Database” and then “Cybersecurity”.

Let us consider an example of using this method to automatically establish connections between informatics concepts (topics) based on students' answers to test questions.

3.2. Building a concept map for the subject of informatics

A concept map is constructed using the Apriori algorithm based on the student's test answers. The algorithm determines which concepts the student failed to learn based on the correct and incorrect answers during the test process. The step-by-step implementation of the algorithm is shown in Fig.1.

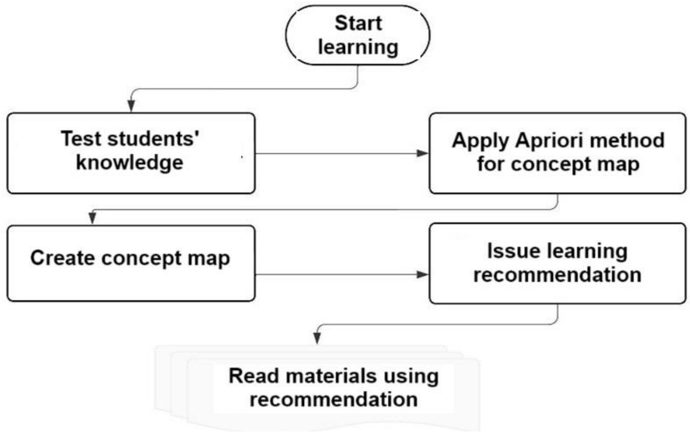

Fig. 1. Algorithm for building a concept map

The concept weight matrix of the questions and the scoring matrix of the test questions answered by the students are described as follows:

$$G = \begin{bmatrix} 0\,1\,0\,0\,1 \\ 1\,0\,0\,0\,1 \\ 0\,0\,1\,0\,0 \\ 1\,0\,0\,1\,1 \\ 0\,1\,0\,0\,0 \end{bmatrix},$$

$$TC = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & 0.7 & 0 & 0.3 & 0 \\ 0 & 0 & 0.8 & 0.2 & 0 \\ 0 & 0 & 0 & 0.6 & 0.4 \end{bmatrix}.$$

The number of students who answered each question incorrectly and correctly based on the students' G matrix is shown in Table 1 and the concept-question TC matrix relationship is shown in Table 2.

Assume that the minimum support value of the questions makes up 40% of the total questions. The support value of 1-element correct and incorrect questions is shown in Table 3. Correct answers get a value of “1”, and incorrect answers get a value of “0”.

Table 1

Students' answers to test questions

Test	Students				
	L_1	L_2	L_3	L_4	L_5
T_1	0	1	0	0	1
T_2	1	0	0	0	1
T_3	0	0	1	0	0
T_4	1	0	0	1	1
T_5	0	1	0	0	0

Table 2

Concept-question relationship

Test	Concepts				
	C_1	C_2	C_3	C_4	C_5
T_1	1	0	0	0	0
T_2	0.5	0.5	0	0	0
T_3	0	0.7	0	0.3	0
T_4	0	0	0.8	0.2	0
T_5	0	0	0	0.6	0.4

Table 3

Support value of 1-element answers

Test	Support value of correct answers	Support value of incorrect answers
T_1	2	3
T_2	2	3
T_3	1	4
T_4	3	2
T_5	1	4

The support value of 2-element correct and incorrect answers to the questions that students answered incorrectly for all question pairs is given in Table 4.

Table 4

Support value of 2-element answers

Test	Support value of correct answers	Support value of incorrect answers	Test	Support value of correct answers	Support value of incorrect answers
$T_1 \rightarrow T_2$	1	2	$T_3 \rightarrow T_4$	0	1
$T_1 \rightarrow T_3$	0	2	$T_3 \rightarrow T_5$	0	3
$T_1 \rightarrow T_4$	1	1	$T_4 \rightarrow T_1$	1	1
$T_1 \rightarrow T_5$	1	3	$T_4 \rightarrow T_2$	2	2
$T_2 \rightarrow T_1$	1	2	$T_4 \rightarrow T_3$	0	1
$T_2 \rightarrow T_3$	0	2	$T_4 \rightarrow T_5$	0	1
$T_2 \rightarrow T_4$	2	2	$T_5 \rightarrow T_1$	1	3
$T_2 \rightarrow T_5$	0	2	$T_5 \rightarrow T_2$	0	2
$T_3 \rightarrow T_1$	0	2	$T_5 \rightarrow T_3$	0	3
$T_3 \rightarrow T_2$	0	2	$T_5 \rightarrow T_4$	0	1

Based on the support value of correct and incorrect answers, the confidence is calculated using formula (1), and the degree of relationship between the concepts $C_1 \rightarrow C_2$ for correct answers is calculated using formula (2):

For the $T_1 \rightarrow T_2$ question pair $min(1, 0.5) * 0.5 = 0.25$ (I = 0.5),

For the $T_1 \rightarrow T_3$ question pair $min(1, 0.7) * 0 = 0$ (I = 0).

The maximum value is selected from these relationships to find the relationship degree for the concepts $C_1 \rightarrow C_2$:

$$max(0.25, 0) = 0.25\ (I = 0.5).$$

Note that for all concept pairs, confidence and relationship degrees are calculated based on correct and incorrect answers.

Let us assume that the minimum confidence set by the user is less than 0.75. We delete the relationships that are smaller than this number. Then, for all question pairs, the confidence of students' correct and incorrect answers and the degree of relationship between the concepts for each question in that pair are obtained using formulas (1), (2) and (3) in Table 5.

Table 5

Confidence and relationship degree

Concept	Relationship degree	Confidence
$C_1 \rightarrow C_5$	0.4	1
$C_2 \rightarrow C_5$	0.3	0.75
$C_3 \rightarrow C_1$	0.5	1
$C_3 \rightarrow C_2$	0.5	1
$C_4 \rightarrow C_1$	0.45	0.75
$C_4 \rightarrow C_5$	0.225	0.75

Based on the concept relationship table, the concept map for the concepts of informatics will be as shown in Fig. 2. From the concept map obtained in this way, it can be seen that concept C_5 cannot be learned before concepts C_1, C_2, C_4, and that concepts C_1 and C_2 must be learned before concept C_3.

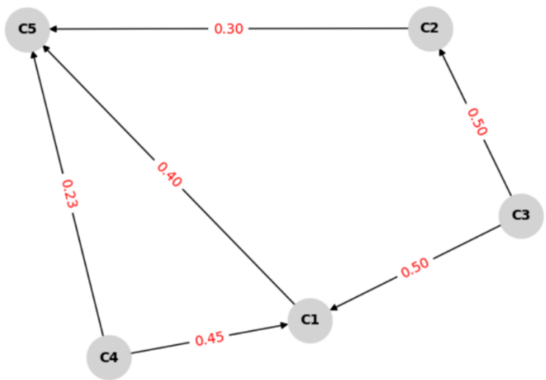

Fig. 2. Concept map

Based on the created concept map, the student is advised on the way to study the topics. This, in turn, ensures that the student has a better grasp of the knowledge and saves time when re-studying the topics that they failed to learn.

It should be noted that the Python programming language "numpy", "pandas", "matplotlib" and "networkx" libraries were used to conduct computer experiments.

4. Conclusion

The article presents the procedure for applying a method for automatically building a concept map in teaching informatics. This method provides teachers with a quickly automatically built concept map to identify concepts (topics) that students have difficulty learning. It is a structured form of knowledge to better understand the topics of the subject of informatics, facilitate learning, and solve complex problems. The use of concept maps ensures that the educational environment is more interactive and productive.

The proposed method builds a concept map based on both incorrect and correct answers based on association rules.

References

[1] K.R. Ayda-Zade, A.E. Aliyeva, Analysis of the problems of teaching computer science in higher education institutions, Transaction of Azerbaijan National Academy of Sciences, Series of Physical-Technical and Mathematical Sciences: Informatics and Control Problems, 38 No.3 (2018) pp.78-84. (in Azerbaijani)

[2] J.D. Novak, D.B. Gowin, Learning how to learn, Book. Cambridge University Press, (1984), doi:10.1017/CBO9781139173469

[3] D.Venkatesan, C. Bharathi Priya, Concept map based learning and evaluation – a review, International Journal of Civil Engineering and Technology (IJCIET). 8 No.8 (2017) pp.947-954.

[4] V. Devedzic, J. Jovanovic, Developing open and interoperable learning environments using concept maps, Educational Technology & Society. 18 No.3 (2015) pp.56-68.

[5] Y. Liu, J. Liang, Y. Pan, Using fuzzy cognitive maps to model adaptive learning in intelligent tutoring systems, Applied Soft Computing. 76 (2019) pp.654-666.

[6] C. Liu, Y. Wang, L. Wang, Personalized concept map generation for adaptive online learning, Computers & Education. 108 (2017) pp.163-173.

[7] H. Buitrago, J. M., & Rios, C. E. P, An adaptive learning approach based on concept mapping and ontologies, Procedia Computer Science. 138 (2018) pp.629-636.

[8] R. Nkambou, J. Bourdeau, R. Mizoguchi, (Eds.), Book. Advances in Intelligent Tutoring Systems, Springer, (2010), doi:10.1007/978-3-642-14363-2

[9] S.C. Lima, E.C. Silva, I.I. Bittencourt, Automatic concept map generation for learning environments, Journal of Intelligent & Fuzzy Systems. 33 No.5 (2017) pp.3145-3156.

[10] V. Devedzic, J. Jovanovic, Developing open and interoperable learning environments sing concept maps, Educational Technology & Society. 18 No.3 (2015) pp.56-68.

[11] C.H. Lee et al., Application of automatically constructed concept map of learning to conceptual diagnosis of e-learning, Expert Systems with Applications. 36 (2009) pp.1675-1684.

[12] S.M. Bai, S.M. Chen, Evaluating students' learning achievement using fuzzy membership functions and fuzzy rules, Expert Systems with Applications. 34 No.1 (2008a) pp.339-410.

[13] S.M. Bai, S.M. Chen, Automatically constructing concept maps based on fuzzy rules for adaptive learning systems, Expert Systems with Applications. 35 No.1 (2008b) pp.41-49.

[14] R. Agrawal, R. Srikant, Fast algorithms for mining association rules, In Proceedings of the 20th international conference on very large database, Santiago, Chile. (1994) pp.487-499.

[15] C. Shyi-Ming, B. Shih-Ming, Using data mining techniques to automatically construct concept maps for adaptive learning system, Expert Systems with Applications. 37 No.6 (2010) pp.4496-4503.