Paper presented at the 23rd international scientific and professional conference "Technical education as a factor of life-long education," University of Matej Bela, Banska Bystrica, September 3-4, 2007.

Refererence:

Lukaš, M., Lešković, D. (2007). Educational system efficiency improvement using knowledge discovery in databases. In J. Pavlovkin (Ed.), *Tehnické vzdelávanie v preocese zmien: Vedecká monografia z prispevkov 23. medzinárodnej vedecko-odbornej konferenie Technické vzdelávanie ako súčast' všeobecného vzdelávania* [Technical education in the process of change: Scientific monograph with papers from 23rd international scientific and professional conference Technical education as a factor of lifelong education] (pp. 179-188). Banska Bystrica, Slovakia: Katedra techniky a technológií, Fakulta prírodných vied, Univerzita Mateja Bela v Banskey Bystrici.

EDUCATIONAL SYSTEM EFFICIENCY IMPROVEMENT USING KNOWLEDGE DISCOVERY IN DATABASES

Mirko LUKAŠ – Darko LEŠKOVIĆ

Sveučilište J. J. Strossmayera, Filozofski fakultet Osijek, Croatia

Abstract

This study describes one of possible way of usage ICT in education system. We basically treated educational system like Business Company and develop appropriate model for clustering of student population. Modern educational systems are forced to extract the most necessary and purposeful information from a large amount of available data. Clustering (segmentation) allows arranging objects into groups and is especially suitable for discovering personal students' characteristics At the end, educational system will be more effectiveness after implementation stabile ICT, and appropriate usage of them. In this study we recognized some benefits as possible positive results and changes within the education system.

1. Introduction

Educational systems are today at a turning point at all levels. In fact, the educational system tends to educate students born into the digital world trying to teach in the way these students comprehend the world. This often means using information in digital format. Should this trend continue, within the next five years there will be significant changes in education that will affect teaching, learning, research and administration: e-learning (Online Education), Electronic Invoicing, and Advanced Information Systems such as ERP (Enterprise Resource Planning). In addition, new problems are emerging, such as data security. All these elements will result in rapid transformation of education systems into virtual classrooms, and thereby certainly of the systems' methods[12]. It is clear that higher education systems are most suitable for rapid implementation of the above mentioned, but the changes will also reach the lower education systems, eventually.

Almost every person around the world has access to a vast amount of information due to the Internet and globalization, which by themselves do not imply knowledge. Information should be transformed into knowledge in order to make it useful; otherwise, it remains dead and useless assets. Therefore, it is a challenge for education systems to offer a "magic formula" for rapid and efficient

[12] M.Zastrocky, M.Harris, B.Rust, J.M.Lowendahl, K.Bell: IT is transforming education, Gartner research No. G00144766, 30.11.2006.

transformation of information into knowledge. This is the "competitive advantage" of educational systems in relation to broadly accessible information. At the same time, education systems compete amongst themselves—public schools, private schools, business schools—offering information. They differ only in efficiency in transforming information into knowledge which becomes practical student's knowledge. These are more than sufficient reasons that make us observe educational systems the way we observe modern companies as efficient systems in market economy.

For this purpose it will be necessary to apply "educational marketing" in order to inform the "potential students" that they will not be able to obtain the useful knowledge themselves without additional education. Development of marketing concept is today in the phase of intensive development of personalized relationships with customers, with the purpose of finding and keeping the most profitable ones[13]. Developing the "technology of relationships with customer" the market leaders transform the series of random transactions into relationships.

We may develop relationships only when we know our customers (buyers, employees, suppliers, distributors, students, in a word – partners), when we invest into development of these relationships and learn about their needs.

2. Study goals and objectives

Study objective is to explore the possibility of using knowledge discovery in databases with the purpose of improving complete education system. Knowledge discovery may, for example, enable further improvements to education system:

a) Get knowing students, their capabilities and tendencies better, in order to direct them, help them to choose occupation and plan a career
b) significantly improving the teachers' competence based on improved knowledge of the students and their capabilities
c) raising the profitability of educational system based on investing into the "right personnel"

Various methods of knowledge discovery from a vast amount of transactional data (all students' interactions) are used to find meaningful regularity, i.e., common characteristics of the students (clustering). This is a way to achieve personalized approach to each student who wants to be treated individually.

The goal and the purpose of the study is to describe the improvement of efficiency in educational systems using information and communication technologies, paying attention to knowledge discovery in databases using cluster analysis. The importance of developing individualized relationships with students as an important element of improving efficiency of education system will be analyzed.

---

[13] prema: Panian, Ž Izazovi elektroničkog poslovanja, Narodne novine Zagreb, 2002.

3. Recognition of students needs

Information and Communication Technology (ICT) has become the theme people around the world talk most often. ICT represents the combination of telecommunications (distance communication) and information (information science, computer science), and offers great possibilities of its use in business activity in general[14]. In that sense the business efficiency also relies on ICT, and it is obvious that it has to be observed through the prism of ICT.

The authors, such as J. Ridderstrale and K. A. Nordstrom state that ICT is obviously significant, but the technology itself will not create efficient systems. In most systems the technology has become like the air we breathe, the water we drink, the sewage system, toilet, water-supply, electricity.... It is available to all the "players" i.e., participants[15]. So, it is necessary, but not sufficient for creating and maintaining efficient systems. The solution may be found in people: teachers, employees, expert assistants, social partners, students, and in creating and developing individual relationships.

Thornton A. May[16] said: „Technology doesn't make you less stupid; it just makes you stupid faster. Basically, we have Star Wars technology, factory-level deployment, and sit-around-the-campfire human behavior".

In order to create relationships with students it is necessary to recognize and satisfy their needs. First, let's see what the need and the desire are. P. Kotler defines the need as the state of lacking some basic human requirements such as food, clothing, shelter, safety, property, self respect, etc, which must be satisfied. The desire is defined as satisfying particular needs. Desires change under the influence of society, i.e., the institutions such as church, school, family, organizations. The desires greatly depend on the person's environment. An average European has different desires than an average person from central Africa. Also, the desires greatly depend on social status and are conditioned by habits, cognition, etc.

Finally, there is a demand. The demand represents a desire for a particular product or a service which can be fulfilled and which the person can afford. Desires become demands when they are supported by purchasing power and free will. Therefore it is essential to recognize not only the wishes, wants, or desires, but also the demands for a particular product/service. Namely, it will not be enough merely to recognize the desire, if it is not supported by the will and the ability to buy[17].

---

[14] Srića, V.; Muller, J. Put k elektroničkom poslovanju, Sinergija, Zagreb, 2001.

[15] Ridderstrale J., Nordstrom, K.A., „Karaoke kapitalizam, management za čovječanstvo", Differo, 2004.

[16] May, A.T., Fast Company magazine, ožujak 2002.

[17] Kotler, P. Upravljanje marketingom, analiza, planiranje, primjena i kontrola, 9. izdanje, Mate Zagreb, 2001.

The demand stimulation will be efficient if a product or a service is presented, i.e., made suitable, attractive, acceptable and available, so the targeted students are able to satisfy their needs that resulted from desire. Educational systems affect our desires but they do not produce the needs. The needs, namely, precede the marketing conception of educational systems which they do not have or have not formed yet.

The students' needs and desires are not always easy to recognize. The students are often unaware of their needs or are unable to express them adequately. For instance, there is an unusual survey result on choosing an occupation with elementary school students from the same class: all students answered that they want to be journalists. Such a "uniform" commitment shows an obvious lack of maturity, i.e., the great influence of external effects on them. It is not impossible that all students from the same class have similar or identical tendencies[18].

So, the survey on students' occupation selection did not prove to be efficient method. More efficient is the method of knowledge discovery in the student's database. Knowledge discovery methods may extract information on students' interests, capabilities, skills, and tendencies, and it is more efficient to choose the occupation using this method[19].

Modern education systems should recognize all sorts of student's needs in order to react correctly and to recognize them:

a) Expressed needs (for instance, lower school fees)
b) Actual needs (for instance, besides the school fee, there is also the possibility of finding employment after finishing school)
c) Unexpressed needs (for instance, expecting to have good relationship with the teacher)
d) Satisfactory needs (for instance, free electronic materials, e-mail consultations, additional activities)
e) Hidden needs (for instance, wanting to achieve a status which is a result of a particular education level)

Education systems that wish to precede the competition should not be satisfied by reacting to the common requirements and desires of anonymous students. They must create their own data warehouses and use them for data mining in order to learn more about each of their students and find out what are their actual needs (knowledge discovery). The lower grades elementary school teachers are able to know almost everything about their students, because their memory is sufficient for the task. On the other hand large educational systems cannot rely upon the memory of their employees; they have to create integrated data warehouses that would include all student transactions, in order to create new knowledge from them.

[18] http://mrav.ffzg.hr/zanimanja/

[19] Luan, J. Data Mining Applications in Higher Education, SSPS Inc. 2004; http://www.ssps.com/ 09.08.2007.

Information from educational data warehouse enables not only reacting to students' demands, but also recognizes actual, unexpressed and hidden needs of the students. In this connection is a conclusion of G. Hamel and C.K. Prahalad from 1994, that it is known that the customers lack foresight. Namely, the organizations should be foresighted, and create products (services, offers, solutions) that will satisfy actual customer demand[20].

4. Knowledge discovery in databases

Knowledge discovery in databases is defined as "the non-trivial extraction of implicit, unknown, and potentially useful information from data"[21]. Cooperating mutually, the process of knowledge discovery takes raw data from data mining, and carefully transforms them into useful and understandable information[22]. Data mining is the process of extracting trends or patterns from data.

Techniques of knowledge discovery in databases share following characteristics:

a) All approaches deal with large amounts of data
b) Efficiency is required, due to volume of data
c) Accuracy is an essential element
d) All approaches use some form of automated learning
e) All produce some interesting results

Development of information and communication systems resulted with relatively easy and cheap system of storing data into databases, what brings questions like: may the historic data in databases be used to develop process models that would serve to generate hidden data; can the developed process models contribute to the analysis of past development of the system or the subsystems and produce concrete results; can future development of the system be predicted based on the process models within a specified period of time. Expanded database usage and new dynamic data exploration approach facilitate obtaining hidden information from large data sets that are significant for obtaining new information, discovering knowledge based on the data, and developing new capital value.

---

[20] Hamel, G.; Prahalad, C.K. Seeing the futute first, Fortune, 05.09.1994.
[21] Frawley, W.J., Piatetsky-Shapiro, G., and Matheus, C. Knowledge Discovery In Databases: An Overview, AAAI Press/MIT Press, Cambridge, MA., 1991.
[22] Fayyad, U.M., Piatetsky-Shapiro, G., and Smyth, P. From Data Mining To Knowledge Discovery, AAAI Press/The MIT Press, Menlo Park, CA., 1996.

Chart 1 Four revolutionary steps in data analysis[23]

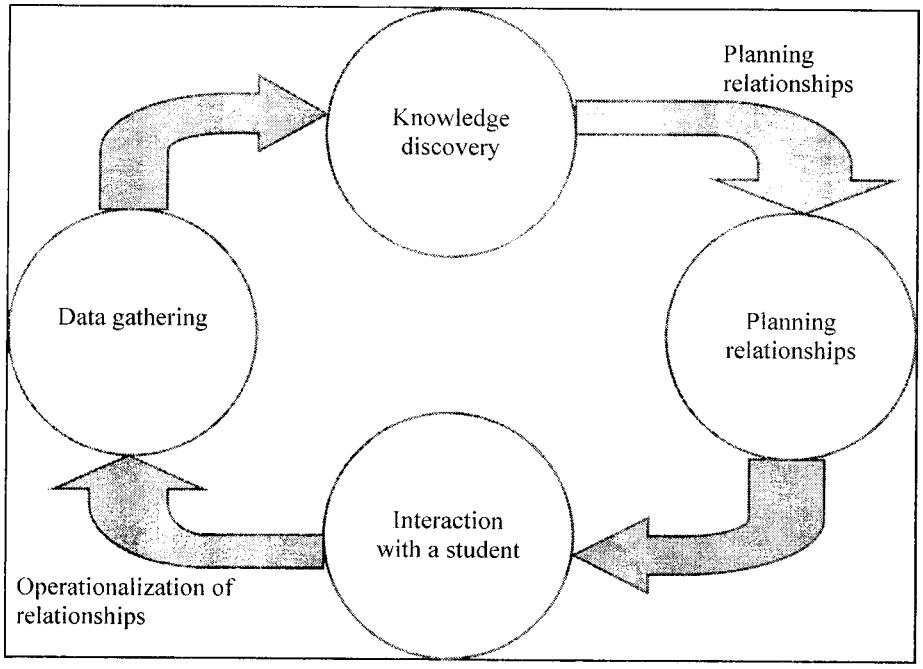
Period	Evolution steps	Commercial (private sector) questions	Technology	Characteristics
1960s	Data collection	What is the company's revenue?	Computers, tapes, disks	Static delivery of historic data
1980s	Data Access	What was the sales realization within a sector last month?	Relation databases, SQL, ODBC	Dynamic delivery of historic data at one level
1990s	Data warehousing and Decision Support system	What was the sales realization for a particular product in the particular sector last month? Drill down the Virovitica region!	OLAP, multidimensional databases, data warehousing	Multilevel dynamic delivery of historic data
Today	Data Mining and Knowledge Discovering	What can happen to the sales realization in Virovitica region next month and why?	Advanced algorithms, multiprocessor computers, vast databases	Predictable and proactive delivery of information

4.1. Knowledge discovery process

Knowledge discovery process can be followed through few basic steps: data selection, data purification, inclusion of well-known a priori knowledge and the correct interpretation of results of data mining process.
Knowledge discovery steps may be defined and described as follows:

a) Data selection – the first step or a phase is to select target group of data that will be used in knowledge discovery process. These may be the information about students, age limits, knowledge, success, capabilities, and alike.
b) Data purification – in this phase the data is accessed from various computers and databases, then purified and matched.
c) Data reduction and projection – at this step the data from transaction databases and other sources are transformed into multidimensional bases. For instance, dimension base of secondary school students consists of time dimension, students, teachers, subjects, interests, capabilities, and similar.
d) Determining the best data mining method – the last step serves to choose the best data mining method, for instance, classification, clustering, market basket analysis, and similar.
e) Finally, there is a correct interpretation followed by making conclusions (norms) as a result of knowledge discovering process.

[23] Ljubetić, V. Upravljanje znanjem primjenom alata poslovne inteligencije, magistarski rad, EFZG 2005.



Pictures 1 Knowledge discovering process for improving students relationship management[24]

4.2. Knowledge discovery methods

Numerous methods of knowledge discovery from data are known today: cluster analysis, neural networks, decision trees, factor analysis. For the study purposes we will describe cluster analysis by the K-means method. K-means allows arranging objects into groups and is especially suitable for discovering personal students' characteristics.

4.2.1 Cluster analysis

Cluster analysis (data clustering, taxonomy analysis) is the basic method for knowledge discovery from data, which is used to classify objects into different groups or subgroups (clusters) that satisfy two main criteria:

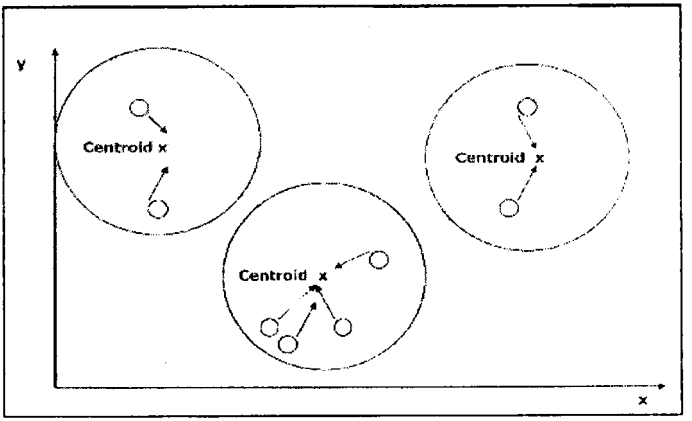
a) each group is homogenous (examples that belong to the same group are similar to each other)

[24] Bona A., Radcliffe J. Eight Building Blocks of CRM: A Framework for Success, Gartner predavanje, Ljubljana, lipanj 2002.

b) each group should be different from other groups (examples that belong to one group are significantly different from the examples of other groups).

The goal of cluster analysis is to merge the objects into clusters based on the similarity of the objects. Similarity is a predefined criterion calculated from the objects' observation (measuring).

K-means algorithm has as an input a predefined number of clusters - K-value. Mean algorithm value refers to the "average" location in multidimensional space defined by attributes. The value of each attribute of an object represents a distance of the object from the origin of that space along the attribute coordinates. In order to use this geometry efficiently, attribute values must be numeric (nominal attribute values must be transformed into numeric values!), and then normalized, in order to allow fair computation along all coordinates (attributes) in a space.



Pictures 2 K-means algorithm

K-mean algorithm is a simple, iterative procedure in which the concept of centroid plays the central role. Centroid is an artificial point in the space of the object, which represents an average location of the particular group of objects. The coordinates of this point are averages of coordinates of all objects that belong to the group.

This iterative procedure of redefining centroids and distributing objects into corresponding groups usually needs only few iterations to converge satisfactorily.

Most cluster analysis methods use the Euclidian distance formula for measuring the distance within the object (square root of the sum of the squares of distances along each coordinate – space attributes). It is necessary to first transform and standardize the nominal attributes. The importance of the attributes in the clustering process largely depends upon this transformation. They may be dominant, but also completely irrelevant if transformed the certain way. If the

number of "k" groups (clusters) in the k-means method is not chosen correctly, final results will not be good. The proper way to select the number of groups is to experiment with different number of groups. Clustering technique is used in cases when "natural" grouping of objects is expected in the data. These segments or groups of data should represent groups of objects that have a lot in common. Creating groups of objects prior to application of some other data modeling technique (neural networks, decision trees) may significantly reduce the complexity of the problem by dividing the group of modeling objects. These subgroups of learning objects can then be modeled separately, and such two step procedure might at the end produce improved results (predictive or descriptive).

4.2.2. Using methods in educational system

Using cluster analysis as one of possible methods for discovering knowledge from database in educational system, we may contribute to a series of positive changes and improvements. Following benefits may be recognized as possible positive results and changes within the education system: a) get knowing the capabilities, tendencies and needs of students, b) directing students in choosing occupation and in planning career, c) improving teachers' competence, d) planning student enrolment quotas, e) planning additional and extracurricular activities in order to obtain a recognizable school image, f) cutting education costs, g) recognizing dropouts in time, before they drop out, as well as the students who are not satisfied with the chosen curriculum, h) recognizing students who fluctuate between educational programs.

The stated benefits facilitate, creating following examples of clusters of students[25]:

a) persistent, b) dropout, c) transfer oriented, d) vocational education directed, e) basic skills upgrades, f) students with mixed outcomes, g) transfer speeders – students who quickly accumulated units, h) college historians – students who took classes for a considerable length of time, i) fence sitters, j) skill up graders, k) speeders, l) laggards, m) stop outs – students who left school and later return...

5. Conclusion

During historical development of educational systems their sources of information have been changing. Modern educational systems are forced to extract the most necessary and purposeful information from a large amount of available data. The Internet, as a modern medium, offers (overly) vast amount of information, and the primary task of educational systems is to extract from offered information those that represent new knowledge. The most appropriate method for knowledge discovery in databases from our analysis is the cluster analysis, which

[25] Luan, J. Data Mining Applications in Higher Education, SSPS Inc. 2004; http://www.ssps.com/ 09.08.2007.

enables forming groups of objects with shared characteristics. Shared characteristics are specific also to all participants and education factors in the didactical triangle. This study proved that cluster analysis enables targeted forming of groups of students with shared characteristics, what at large contributes to the improvement of complete educational system.

Table of authors

1. Bona A., Radcliffe J. Eight Building Blocks of CRM: A Framework for Success, Gartner predavanje, Ljubljana, June 2002.
2. Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P. From Data Mining To Knowledge Discovery, AAAI Press/The MIT Press, Menlo Park, CA., 1996.
3. Frawley, W.J., Piatetsky-Shapiro, G., and Matheus, C. Knowledge Discovery In Databases: An Overview, AAAI Press/MIT Press, Cambridge, MA., 1991.
4. Hamel, G., Prahalad, C.K. Seeing the futute first, Fortune, 05.09.1994.
5. Kotler, P. Upravljanje marketingom, analiza, planiranje, primjena i kontrola, 9. edition, Mate Zagreb, 2001.
6. Luan, J. Data Mining Applications in Higher Education, SSPS Inc. 2004;
7. Luan, J. Data Mining Applications in Higher Education, SSPS Inc. 2004;
8. Ljubetić, V. Upravljanje znanjem primjenom alata poslovne inteligencije, MSc thesis, Ekonomski fakultet Zagreb 2005.
9. May, A.T., Fast Company magazine, mart 2002.
10. Panian, Ž Izazovi elektroničkog poslovanja, Narodne novine Zagreb, 2002.
11. Ridderstrale J., Nordstrom, K.A., Karaoke kapitalizam, management za čovječanstvo, Differo, 2004.
12. Srića, V.; Muller, J. Put k elektroničkom poslovanju, Sinergija, Zagreb, 2001.
13. Zastrocky, M., Harris, M., Rust, B., Lowendahl, J.M., Bell, K.: IT is transforming education, Gartner research No. G00144766, 30.11.2006.
14. http://mrav.ffzg.hr/zanimanja/ 10.08.2007.
15. http://www.ssps.com/ 09.08.2007.
16. http://www.ssps.com/ 09.08.2007.