UDC 004.047

EDUCATIONAL DATA MINING – PHASES AND TECHNICAL METHODS

ARTSIOM RODCHANKAU, ARKADIY OSKIN
Polotsk State University, Belarus

The article deals with Educational data mining aims and objectives. The main phases in EDM are defined and described. Popular data mining methods are considered. The article concludes with an analysis of the current state of affairs in this research area.

INTRODUCTION. Data mining, also known as Knowledge Discovery in Databases (KDD), is used to find out valid, novel, and potentially useful information from large amount of data. Data mining has been applied in a great number of fields, including retail sales, bioinformatics, and counter-terrorism. In recent years, there has been an increasing interest in data mining application to investigate scientific questions within educational research. This research field has been termed educational data mining. Educational data mining (also referred to as "EDM") is defined as a discipline concerned with developing methods for exploring the unique and increasingly large-scale data that come from educational settings (e.g. universities and intelligent tutoring systems), and using those methods to better understand students, and the settings which they learn in.

EDM refers to techniques and tools designed for automatically extracting meaning from large repositories of data generated by people's learning activities in educational settings. At a high level, the field seeks to develop and improve methods for exploring this data, which often has multiple levels of meaningful hierarchy, in order to discover new insights about how people learn in the context of such settings. The main objective of applying Data Mining to educational data is to analyze educational Data contents, models, to summarize/analyze the learner's discussions, etc. Education Data Mining concentrates on the computing process models which focus on Education context (e.g. assessments, attendance and progress). In educational system, a student's performance is determined by the term work, attendance and end semester examination. The term work is carried out by the teacher based upon a student's performance in educational activities such as class tests, assignments, attendance [1].

Phases in Educational Data Mining

Educational data mining is an emerging *growing* research community. A lot of mining techniques has been accomplished in a variety of educational contexts. The major objective is to translate raw data into meaningful information about the learning process, to make better decisions about the design and trajectory of a learning environment.

EDM generally consists of four phases [2]:
1. Discover relationships
2. Validate relationships
3. Make Predictions
4. Decision Making

1) Discover Relationships

The first phase in the EDM process (not including preprocessing) is to discover relationships between data. It is accomplished by searching in a repository of data in an educational environment with the objective of finding consistent relationships between data variables. Various algorithms are used for finding such relationships: classification, regression, clustering, factor analysis, social network analysis, association rule mining, and sequential pattern mining.

2) Validating Relationships

Discovered relationships must be validated in order to avoid over fitting.

3) Make Predictions

Valid relationships are used to make predictions about further events in the learning environment.

4) Decision Making

Predictions made in the previous phase are used to support decision-making processes and in making policy decisions. In phases 3 and 4, data is often visualized and distillateв to enable a human to quickly identify or classify features of the data. Visualising Data is one of the best practices in research.

Technical Methods

EDM methods fall into the following general categories: prediction, clustering, relationship mining, discovery with models, and distillation of data for human judgment (Table 1) [3].

Table 1. – Basic types of educational data mining methods

Types of Methods	Goals
Prediction	Develop a model which can infer a single aspect of the data (predicted variable) from some combination of other aspects of the data (predictor variables)
Clustering	Find data points that naturally group together, splitting the full data set into a set of categories
Relationship Mining	Discover relationships between variables
Discovery with Models	A model of a phenomenon developed with prediction, clustering, or knowledge engineering is used as a component in further prediction or relationship mining
Distillation of Data for Human Judgment	Data is distilllated to enable a human to quickly identify or classify features of the data

The first three types are largely acknowledged to be universal across types of data mining (albeit in some cases with different names). The fourth and fifth types achieve particular prominence within educational data mining.

Educational data mining researchers view the following as the goals for their research [2]:

1. Predicting students' future learning behavior by creating student models that incorporate such detailed information as students' knowledge, motivation, metacognition, and attitudes;

2. Discovering or improving domain models that characterize the content to be learned and optimal instructional sequences;

3. Studying the effects of different kinds of pedagogical support that can be provided by learning software;

4. Advancing scientific knowledge about learning and learners through building computational models that incorporate models of the student, the domain, and the software pedagogy.

CONCLUSION. **Educational Data Mining** is an emerging discipline concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in. Educational Data Mining focuses on developing new tools and algorithms for discovering data patterns. EDM develops methods and applies techniques from statistics, machine learning, and data mining to analyze data collected during teaching and learning. EDM tests learning theories and informs educational practice.

The application of data mining methods in the educational sector is an interesting phenomenon. Techniques in educational organizations help us to learn a student's performance and behavior, designing courses, curriculum and to motivate students on various parameters.

REFERENCES

1.   Educational data mining — [Electronic resource]. – Mode of access: https://en.wikipedia.org/wiki/Educational_data_mining. – Date of access: 10.02.2018.
2.   Baker, R.S.J.d. The State of Educational Data Mining in 2009 : A Review and Future Visions / R.S.J.d. Baker, K. Yacef.
3.   Baker, R.S.J.d. Data Mining for Education / R.S.J.d. Baker ; Carnegie Mellon University. – Pittsburgh, Pennsylvania, USA.