Data Mining for Education

Ryan S.J.d. Baker, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA
rsbaker@cmu.edu

Article to appear as
Baker, R.S.J.d. (in press) Data Mining for Education. To appear in McGaw, B., Peterson, P., Baker, E. (Eds.) *International Encyclopedia of Education (3rd edition)*. Oxford, UK: Elsevier.

This is a pre-print draft. Final article may involve minor changes and different formatting.

I would like to thank Cristobal Romero, Sandip Sinharay, and Joseph Beck for their comments and suggestions on this document, and Joseph Beck and Jack Mostow for their permission to discuss their research as a "best practices" case study in this article.

Data Mining for Education

Ryan S.J.d. Baker, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

Introduction

Data mining, also called Knowledge Discovery in Databases (KDD), is the field of discovering novel and potentially useful information from large amounts 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 increasing interest in the use of data mining to investigate scientific questions within educational research, an area of inquiry termed educational data mining. Educational data mining (also referred to as "EDM") is defined as the area of scientific inquiry centered around the development of methods for making discoveries within the unique kinds 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 methods often differ from methods from the broader data mining literature, in explicitly exploiting the multiple levels of meaningful hierarchy in educational data. Methods from the psychometrics literature are often integrated with methods from the machine learning and data mining literatures to achieve this goal.

For example, in mining data about how students choose to use educational software, it may be worthwhile to simultaneously consider data at the keystroke level, answer level, session level, student level, classroom level, and school level. Issues of time, sequence, and context also play important roles in the study of educational data.

Educational data mining has emerged as an independent research area in recent years, culminating in 2008 with the establishment of the annual International Conference on Educational Data Mining, and the Journal of Educational Data Mining.

Advantages Relative to Traditional Educational Research Paradigms

Educational data mining offers several advantages, vis-à-vis more traditional educational research paradigms, such as laboratory experiments, in-vivo experiments, and design research.

In particular, the advent of public educational data repositories such as the PSLC DataShop and the National Center for Education Statistics (NCES) data sets has created a base which makes educational data mining highly feasible. In particular, the data from these repositories is often both ecologically valid (inasmuch as it is data about the performance and learning of genuine students, in genuine educational settings, involved in authentic learning tasks), and increasingly easy to rapidly access and begin research with. Balancing feasibility with ecological validity is often a difficult challenge for researchers in other educational research paradigms. By contrast, researchers who use data from these repositories can dispense with traditionally time-consuming steps such as subject recruitment (e.g. recruitment of schools, teachers, and students), scheduling of studies, and data entry (since data is already online). While the use of previously collected data has the potential to limit analyses to questions involving the types of data collected, in practice data from repositories or prior research has been useful for analyzing research questions far outside the purview of what the data were originally intended to study, particularly given the advent of models that can infer student attributes (such as strategic behavior and motivation) from the type of data in these repositories.

This increase in speed and feasibility has had the benefit of making replication much more feasible. Once a construct of educational interest (such as off-task behavior, or whether or not a skill is known) has been empirically defined in data, it can be transferred to new data sets. The transfer of constructs is not trivial – often, the same construct can be subtly different at the data level, within data from a different context or system – but transfer learning and rapid labeling methods have been successful in speeding up the process of developing or validating a model for a new context. This has led to many educational data mining analyses being replicated across data from several learning systems or contexts.

Increasingly, the existence of data from thousands of students, having broadly similar learning experiences (such as using the same learning software), but in very different contexts, gives leverage that was never before possible, for studying the influence of contextual factors on learning and learners. It has historically been difficult to study how much the differences between teachers and classroom cohorts influence specific aspects of the learning experience; this sort of analysis becomes much easier with educational data mining. Similarly, the concrete impacts of fairly rare individual differences have been difficult to statistically study with traditional methods (leading case studies to be a dominant research method in this area) – educational data mining has the potential to extend a much wider tool set to the analysis of important questions in individual differences.

Main Approaches

<put Table 1 here>

There are a wide variety of current methods popular within educational data mining. These methods fall into the following general categories: prediction, clustering, relationship mining, discovery with models, and distillation of data for human judgment. The first three categories are largely acknowledged to be universal across types of data mining (albeit in some cases with different names). The fourth and fifth categories achieve particular prominence within educational data mining.

Prediction

In prediction, the goal is to 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). Prediction requires having labels for the output variable for a limited data set, where a label represents some trusted "ground truth" information about the output variable's value in specific cases. In some cases, however, it is important to consider the degree to which these labels may in fact be approximate, or incompletely reliable.

Prediction has two key uses within educational data mining. In some cases, prediction methods can be used to study what features of a model are important for prediction, giving information about the underlying construct. This is a common approach in programs of research that attempt to predict student educational outcomes (cf. Romero et al, 2008) without predicting intermediate or mediating factors first. In a second type of usage, prediction methods are used in order to predict what the output value would be in contexts where it is not desirable to directly obtain a label for that construct (for example, in previously collected repository data, where desired labeled data may not be available, or in contexts where obtaining labels could change the behavior being labeled, such as modeling affective states, where self-report, video, and observational methods all present risks of altering the construct being studied).

For example, consider research attempting to study the relationship between learning and gaming the system, attempting to succeed in an interactive learning environment by exploiting properties of the system rather than by learning the material. If a researcher has the goal of studying this construct across a full year of software usage within multiple schools, it may not be tractable to directly assess, using non data-mining methods, whether each student is gaming, at each point in time. Baker et al (2008) developed a prediction model by using observational methods to label a small data set, developing a prediction model using automatically collected data from interactions between students and the software for predictor variables, and then validating the model's accuracy when generalized to additional students and contexts. They were then able to study their research question in the context of the full data set.

Broadly, there are three types of prediction: classification, regression, and density estimation. In classification, the predicted variable is a binary or categorical variable. Some popular classification methods include decision trees, logistic regression (for binary predictions), and support vector machines. In regression, the predicted variable is a continuous variable. Some popular regression methods within educational data mining include linear regression, neural networks, and support vector machine regression. In density estimation, the predicted variable is a probability density function. Density estimators can be based on a variety of kernel functions, including Gaussian functions. For each type of prediction, the input variables can be either categorical or continuous; different prediction methods are more effective, depending on the type of input variables used.

Popular methods for assessing the goodness of a predictor include linear correlation, Cohen's Kappa, and A' (the area under the receiver-operating curve – e.g. Bradley, 2007). Percent accuracy is generally not preferred for classification, as values of accuracy are highly dependent on the base rates of different classes (and hence, a very high accuracy can in some cases be achieved by a classifier that simply always predicts the majority class). When computing the goodness of a predictor, it is important to account for non-independence of different observations involving the same student – to achieve this goal, educational data mining researchers often apply meta-analytical methods that can account for partial non-independence, such as Strube's (1985) Adjusted Z, or select overly conservative estimators that assume complete non-independence.

Clustering

In clustering, the goal is to find data points that naturally group together, splitting the full data set into a set of clusters. Clustering is particularly useful in cases where the most common categories within the data set are not known in advance. If a set of clusters is optimal, within a category, each data point will in general be more similar to the other data points in that cluster than data points in other clusters. Clusters can be created at several different possible grain-sizes: for example, schools could be clustered together (to investigate similarities and differences between schools), students could be clustered together (to investigate similarities and differences between students), or student actions could be clustered together (to investigate patterns of behavior) (cf. Amershi & Conati, 2006; Beal, Qu, & Lee, 2006).

Clustering algorithms can either start with no prior hypotheses about clusters in the data (such as the k-means algorithm with randomized restart), or start from a specific hypothesis, possibly generated in prior research with a different data set (using the Expectation Maximization algorithm to iterate towards a cluster hypothesis for the new data set). A clustering algorithm can postulate that each data point must belong to exactly one cluster (such as in the k-means algorithm), or can postulate that some points may belong to more than one cluster or to no clusters (such as in Gaussian Mixture Models).

The goodness of a set of clusters is usually assessed with reference to how well the set of clusters fits the data, relative to how much fit might be expected solely by chance given the number of clusters, using statistical metrics such as the Bayesian Information Criterion.

Relationship Mining

In relationship mining, the goal is to discover relationships between variables, in a data set with a large number of variables. This may take the form of attempting to find out which variables are most strongly associated with a single variable of particular interest, or may take the form of attempting to discover which relationships between any two variables are strongest.

Broadly, there are four types of relationship mining: association rule mining, correlation mining, sequential pattern mining, and causal data mining. In association rule mining, the goal is to find if-then rules of the form that if some set of variable values is found, another variable will generally have a specific value. For example, a rule might be found of the form {student is frustrated, student has stronger goal of learning than goal of performance} → {student frequently asks for help}. In correlation mining, the goal is to find (positive or negative) linear correlations between variables. In sequential pattern mining, the goal is to find temporal associations between events – for example, to determine what path of student behaviors leads to an eventual learning event of interest. In causal data mining, the goal is to find whether one event (or observed construct) was the cause of another event (or observed construct), either by analyzing the covariance of the two events (e.g. TETRAD – Scheines et al, 1994) or by using information about how one of the events was triggered. For example, if a pedagogical event is randomly chosen using automated experimentation (Mostow, 2008), and frequently leads to a positive learning outcome, a causal relationship can be inferred.

Relationships found through relationship mining must satisfy two criteria: statistical significance, and interestingness. Statistical significance is generally assessed through standard statistical tests, such as F-tests. Because large numbers of tests are conducted, it is necessary to control for finding relationships through chance. One method for doing this is to use post-hoc statistical methods or adjustments which control for the number of tests conducted, such as the Bonferroni adjustment. This method can increase confidence that an individual relationship found was not likely to be due to chance. An alternate method is to assess the overall probability of the pattern of results found, using Monte Carlo methods. This method assesses how likely it is that the overall pattern of results arose due to chance.

The interestingness of each finding is assessed in order to reduce the set of rules/ correlations/ causal relationships communicated to the data miner. In very large data sets, hundreds of thousands of significant relationships may be found. Interestingness measures attempt to determine which findings are the most distinctive and well-supported by the data, in some cases also attempting to prune overly similar findings. There are a wide variety of interestingness measures, including support, confidence, conviction, lift, leverage, coverage, correlation, and cosine. Some investigations have suggested that lift and cosine may be particularly relevant within educational data (Merceron & Yacef, 2008).

Discovery with Models

In discovery with a model, a model of a phenomenon is developed via prediction, clustering, or in some cases knowledge engineering (within knowledge engineering, the model is developed using human reasoning rather than automated methods). This model is then used as a component in another analysis, such as prediction or relationship mining.

In the prediction case, the created model's predictions are used as predictor variables in predicting a new variable. For instance, analyses of complex constructs such as gaming the system within online learning have generally depended on assessments of the probability that the student knows the current knowledge component being learned (Baker et al, 2008; Walonoski & Heffernan, 2006). These assessments of student knowledge have in turn depended on models of the knowledge components in a domain, generally expressed as a mapping between exercises within the learning software and knowledge components.

In the relationship mining case, the relationships between the created model's predictions and additional variables are studied. This can enable a researcher to study the relationship between a complex latent construct and a wide variety of observable constructs.

Often, discovery with models leverages the validated generalization of a prediction model across contexts. For instance, Baker (2007) used predictions of gaming the system across a full year of educational software data to study whether state or trait factors were better predictors of how much a student would game the system. Generalization in this fashion relies upon appropriate validation that the model accurately generalizes across contexts.

Distillation of Data for Human Judgment

Another area of interest within educational data mining is the distillation of data for human judgment. In some cases, human beings can make inferences about data, when it is presented appropriately, that are beyond the immediate scope of fully automated data mining methods. The methods in this area of educational data mining are information visualization methods – however, the visualizations most commonly used within EDM are often different than those most often used for other information visualization problems (cf. Kay et al, 2006; Hershkovitz & Nachmias, 2008), owing to the specific structure, and the meaning embedded within that structure, often present in educational data.

Data is distilled for human judgment in educational data mining for two key purposes: identification and classification. When data is distilled for identification, data is displayed in ways that enable a human being to easily identify well-known patterns that are nonetheless difficult to formally express. For example, one classic educational data mining visualization is the learning curve, which displays the number of opportunities to practice a skill on the X axis, and displays performance (such as percent correct or time taken to respond) on the Y axis. A curve with a smooth downward progression that is steep at first and gentler later indicates a well-specified knowledge component model. A sudden spike upwards, by contrast, indicates that more than one knowledge component is included in the model (cf. Corbett & Anderson, 1995).

Alternately, data may be distilled for human labeling, to support the later development of a prediction model. In this case, sub-sections of a data set are displayed in visual or text format, and labeled by human coders. These labels are then generally used as the basis for the development of a predictor. This approach has been shown to speed the development of prediction models of complex phenomena such as gaming the system by around 40 times, relative to prior approaches for collecting the necessary data (Baker & de Carvalho, 2008).

Main Applications

There have been a wide number of applications of educational data mining, as reflected throughout this chapter. In this section, four areas of application that have received particular attention within the field are discussed.

One key area of application is in improving student models, models that provide detailed information about a student's characteristics or states, such as knowledge, motivation, meta-cognition, and attitudes. Modeling the individual differences between students, in order to enable software to respond to those individual differences, is a key theme in educational software research. In the last few years educational data mining methods have enabled considerable expansion in the sophistication of student models. In particular, educational data mining methods have enabled researchers to make higher-level inferences about students' behavior, such as when a student is gaming the system, when a student has "slipped" (making an error despite knowing a skill), and when a student is engaging in self-explanation (cf. Shih, Koedinger, & Scheines, 2008). These richer student models have been useful in two fashions. First, these models have increased our ability to predict student knowledge and future performance – incorporating models of guessing and slipping into predictions of student future performance has increased the accuracy of these predictions by up to 48% (Baker, Corbett, & Aleven, 2008). Second, these models have enabled researchers to study what factors lead students to make specific choices in a learning setting, a type of scientific discovery with models discussed below.

A second key area of application is in discovering or improving models of the knowledge structure of the domain. In educational data mining, methods have been created for rapidly discovering accurate domain models directly from data. These methods have generally combined psychometric modeling frameworks with advanced space-searching algorithms, and are generally posed as prediction problems for the purpose of model discovery (for example, attempting to predict whether individual actions will be correct or incorrect, using different domain models, is one common method for developing these models). Barnes, Bitzer, & Vouk (2005) have proposed algorithms for automatically discovering a Q-Matrix from data. Cen, Koedinger, & Junker (2006) proposed algorithms for using codified expert knowledge about differences between items to drive automated search for IRT models. Pavlik et al (2008) has proposed algorithms for finding partial order knowledge structure models (cf. Desmarais, Maluf, & Liu, 1996), by looking at the covariation of individual items.

A third key area of application is in studying the pedagogical support provided by learning software. Modern educational software gives a variety of types of pedagogical support to students. Discovering which pedagogical support is most effective has been a key area of interest for educational data miners. Learning decomposition (Beck & Mostow, 2008), a type of relationship mining, fits exponential learning curves to performance data, relating student success to the amount of each type of pedagogical support a student has received (with a weight for each type of support). The weights indicate how effective each type of pedagogical support is for improving learning. An illustrative example is given in the next section.

A fourth key area of application of educational data mining is for scientific discovery about learning and learners. This takes on several forms. Applying educational data mining to answer questions in any of the three areas previously discussed (e.g. student models, domain models, and pedagogical support) can have broader scientific benefits; for example, the study of pedagogical support may have the long-term potential to enrich theories of scaffolding. Beyond just these three areas, however, there have been many analyses aimed directly towards scientific discovery. Discovery with models is a key method for scientific discovery via educational data mining. Research on studying whether state or trait factors were better predictors of how much a student would game the system (Baker, 2007) is a prominent example of this approach within educational data mining research. Learning decomposition methods are another prominent method for conducting scientific discovery about learning and learners.

Illustrative Example

In this section, a brief case study is discussed, as a concrete “best practices” example of how the educational data mining method of learning decomposition (a type of relationship mining) was used to determine the relative efficacy of different types of learning material presented to students.

In Beck and Mostow (2008), data was obtained from 346 American elementary school students reading 6.9 million words, over the course of a year, while using intelligent tutor software that teaches reading. These words were presented in the form of stories, and students and the software took turns choosing stories (the software’s choice of stories was based on the student’s approximate grade reading level). Beck and Mostow were interested in determining whether re-

reading a story (a popular option for children) is more or less effective at promoting word learning than encountering the same word in a different story. They were also interested in whether there would be individual differences, such that some students would benefit from a different pattern of practice than others.

Beck and Mostow obtained data for each student's performance in reading each story within the software. Reading time was used as a continuous measure of word knowledge; mis-reading and help-requests were also taken into account, reading opportunities where these behaviors occurred were assigned a time of 3.0 seconds (99.9% of word reads were faster than 3.0 seconds). An exponential model of practice was set up, relating response time to the function:

$$time = A * e^{-b\,(W*t_1+t_2)}.$$

In this equation, parameter A represents student performance on the first opportunity to read a given word, parameter b represents the overall speed of learning, e is 2.718, and t1 and t2 represent the number of times the word is read, within two different types of practice. In this case, t1 was defined as the number of times the word was read when re-reading a story and t2 was defined as the number of times the word was read when reading a story for the first time. W is the relative speed gain associated with the two types of practice. If W equals 1, the two types of practice are considered to be equally effective; if W is above 1, opportunities of type t1 are more effective than opportunities of type t2 (and the reverse holds true if W is below 1).

Across the population of students, the median value of W for re-reading obtained by Beck and Mostow was 0.49, suggesting that re-reading a story leads to approximately half as much learning as reading a new story. 95 of the 346 students had a W parameter statistically significantly under 1, whereas only 7 students had a W parameter value statistically significantly over 1, a statistically significant result across the entire class.

Beck and Mostow next used the values of W from the model in a subsequent logistic regression analysis (an example of discovery with models). In this analysis, the learning decomposition model was used to split the population into students who benefitted from re-reading and students who did not benefit from re-reading, and a variety of explanatory variables were tested to see if they explained which students benefitted from re-reading. This analysis determined that students with overall low reading speed who were receiving special needs learning support actually benefitted from re-reading.

Bibliography

Amershi, S., Conati, C. (2006) Automatic Recognition of Learner Groups in Exploratory Learning Environments. *Proceedings of ITS 2006, 8th International Conference on Intelligent Tutoring Systems.*

Baker, R.S.J.d. (2007) Is Gaming the System State-or-Trait? Educational Data Mining Through the Multi-Contextual Application of a Validated Behavioral Model. *Complete On-Line Proceedings of the Workshop on Data Mining for User Modeling at the 11th International Conference on User Modeling 2007*, 76-80.

Baker, R.S.J.d., Corbett, A.T., Aleven, V. (2008) More Accurate Student Modeling Through Contextual Estimation of Slip and Guess Probabilities in Bayesian Knowledge Tracing. *Proceedings of the 9th International Conference on Intelligent Tutoring Systems,* 406-415.

Baker, R.S.J.d., Corbett, A.T., Roll, I., Koedinger, K.R. (2008) Developing a Generalizable Detector of When Students Game the System. *User Modeling and User-Adapted Interaction,*18, 3, 287-314.

Baker, R.S.J.d., de Carvalho, A.M.J.A. (2008). Labeling Student Behavior Faster and More Precisely with Text Replays. *Proceedings of the First International Conference on Educational Data Mining,* 38-47.

Barnes, T., Bitzer, D., Vouk, M. (2005) Experimental Analysis of the Q-Matrix Method in Knowledge Discovery. *Lecture Notes in Computer Science 3488: Foundations of Intelligent Systems,* 603-611.

Beal, C., Qu, L., Lee, H. (2007) Classifying learner engagement through integration of multiple data sources. *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI-2007).*

Beck, J.E., Mostow, J. (2008). How who should practice: Using learning decomposition to evaluate the efficacy of different types of practice for different types of students. *Proceedings of the 9th International Conference on Intelligent Tutoring Systems,* 353-362.

Bradley, A.P. (1997) The Use of the Area Under the ROC Curve in the Evaluation of Machine Learning Algorithms. Pattern Recognition, 30, 1145-1159.

Cen, H., Koedinger, K., Junker, B. (2006) Learning Factors Analysis - A General Method for Cognitive Model Evaluation and Improvement**.** *Proceedings of the 8th International Conference on Intelligent Tutoring Systems*, 12-19.

Corbett, A.T., & Anderson, J.R. (1995). Knowledge Tracing: Modeling the Acquisiton of Procedural Knowledge. *User Modeling and User-Adapted Interaction, 4,* 253-278.

Desmarais, M.C., Maluf, A., Liu, J. (1996) User-expertise modeling with empirically derived probabilistic implication networks. *User Modeling and User-Adapted Interaction, 5,* 3-4, 283-315.

Hershkovitz, A., Nachmias, R. (2008) Developing a Log-Based Motivation Measuring Tool. *Proceedings of the First International Conference on Educational Data Mining,* 226-233.

Kay, J., N. Maisonneuve, K. Yacef and P. Reimann (2006). The Big Five and Visualisations of Team Work Activity. Proceedings of Intelligent Tutoring Systems (ITS06), M. Ikeda, K. D. Ashley & T-W. Chan (eds). Taiwan. Lecture Notes in Computer Science 4053, Springer-Verlag, 197-206.

Mostow, J. (2008). Experience from a Reading Tutor that listens: Evaluation purposes, excuses, and methods. In C. K. Kinzer & L. Verhoeven (Eds.), *Interactive Literacy Education: Facilitating Literacy Environments Through Technology*, pp. 117-148. New York: Lawrence Erlbaum Associates, Taylor & Francis Group.

Mostow, J. and J. Beck (2006, July 6-8*). Refined micro-analysis of fluency gains in a Reading Tutor that listens.* Thirteenth Annual Meeting of the Society for the Scientific Study of Reading, Vancouver, BC, Canada.

Pavlik, P., Cen, H., Wu, L., Koedinger, K. (2008) Using Item-Type Performance Covariance to Improve the Skill Model of an Existing Tutor. *Proceedings of the First International Conference on Educational Data Mining,* 77-86.

Romero, C., Ventura, S., Espejo, P.G., Hervas, C. (2008) Data Mining Algorithms to Classify Students. *Proceedings of the First International Conference on Educational Data Mining,* 8-17.

Merceron, A., Yacef, K. (2008) Interestingness Measures for Association Rules in Educational Data. *Proceedings of the First International Conference on Educational Data Mining,* 57-66.

Scheines, R., Sprites, P., Glymour, C., Meek, C. (1994) *Tetrad II: Tools for Discovery.* Lawrence Erlbaum Associates: Hillsdale, NJ.

Shih, B., Koedinger, K.R., Scheines, R. (2008) A Response-Time Model for Bottom-Out Hints as Worked Examples. *Proceedings of the First International Conference on Educational Data Mining,* 117-126.

Strube, M.J. (1985) Combining and comparing significance levels from nonindependent hypothesis tests. *Psychological Bulletin,* 97, 334-341.

Walonoski, J., Heffernan, N.T. (2006). Detection and Analysis of Off-Task Gaming Behavior in Intelligent Tutoring Systems. *Proceedings of the 8th International Conference on Intelligent Tutoring Systems*, 382-391.

Further Reading

Moore, A. (2005) *Statistical Data Mining Tutorials.* Available online at http://www.autonlab.org/tutorials/ . Retrieved June 27, 2008.

Baker, R.S.J.d., Barnes, T., Beck, J.E. (Eds.) (2008) *Proceedings of the 1st International Conference on Educational Data Mining.*

Baker, R.S.J.d., Beck, J.E., Berendt, B., Kroner, A., Menasalvas, E., Weibelzahl, S. (2007) *Complete On-Line Proceedings of the Workshop on Data Mining for User Modeling, at the 11th International Conference on User Modeling (UM 2007).*

Heiner, C., Heffernan, N.T., Barnes, T., (Eds.) (2007) *Proceedings of the Workshop on Educational Data Mining, at the 13th International Conference on Artificial Intelligence in Education.*

Romero, C., Ventura, S. (2006) *Data Mining in e-Learning.* Southampton, UK: WIT Press.

Romero, C., Ventura, S. (2007) Educational Data Mining: A Survey from 1995 to 2005. *Expert Systems with Applications*, 33(1), 135-146.

Witten, I.H., Frank, E. (2005) *Data Mining: Practical Machine Learning Tools and Techniques: 2nd Edition.* San Francisco: Morgan Kaufmann.

Tables/Illustrations

Category of Method	Goal of Method	Key applications
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)	Detecting student behaviors (e.g. gaming the system, off-task behavior, slipping); Developing domain models; Predicting and understanding student educational outcomes
Clustering	Find data points that naturally group together, splitting the	Discovery of new student behavior patterns;

	full data set into a set of categories	Investigating similarities and differences between schools
Relationship Mining	Discover relationships between variables	Discovery of curricular associations in course sequences; Discovering which pedagogical strategies lead to more effective/robust learning
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.	Discovery of relationships between student behaviors, and student characteristics or contextual variables; Analysis of research question across wide variety of contexts
Distillation of Data for Human Judgment	Data is distilled to enable a human to quickly identify or classify features of the data.	Human identification of patterns in student learning, behavior, or collaboration; Labeling data for use in later development of prediction model

Table 1: The primary categories of educational data mining