Chapter X: Educational Data Mining and Learning Analytics

Ryan Shaun Joazeiro de Baker and Paul Salvador Inventado

Abstract In recent years, two communities have grown around a joint interest in how big data can be exploited to benefit education and the science of learning: Educational Data Mining and Learning Analytics. This article discusses the relationship between these two communities, and the key methods and approaches of educational data mining. The article discusses how these methods emerged in the early days of research in this area, which methods have seen particular interest in the EDM and learning analytics communities, and how this has changed as the field matures and has moved to making significant contributions to both educational research and practice.

X.1 Introduction

In this article, we will discuss a research area/community with close ties to the learning analytics community discussed throughout this book, educational data mining (EDM). This chapter will introduce the EDM community, its methods, ongoing trends in the area, and give some brief thoughts on its relationship to the learning analytics community.

Educational data mining can be seen in two ways; either as a research community or as an area of scientific inquiry. As a research community, EDM can be seen as a sister community to learning analytics. Educational Data Mining first emerged in a workshop series starting in 2005, which became an annual conference in 2008 and spawned a journal in 2009 and a society, the International Educational Data Mining Society, in 2011. A timeline of key events in the formation of the EDM community can be seen in Figure X.1.

As of this writing, the EDM Society has 240 paid members, and the conference has an annual attendance around the same number. Many of the same people attend both EDM and the Learning Analytics and Knowledge (LAK) conference, and the general attitude between the two conferences is one of friendly collaboration and/or friendly competition.

As an area of scientific inquiry, educational data mining is concerned with the analysis of large-scale educational data, with a focus on automated methods. There is considerable thematic overlap between EDM and learning analytics. In particular, both communities share a common interest in data-intensive approaches to education research, and share the goal of enhancing educational practice. At the same time, there are several interesting differences, with one viewpoint on the

R. S. J. d. Baker - P.S. Inventado

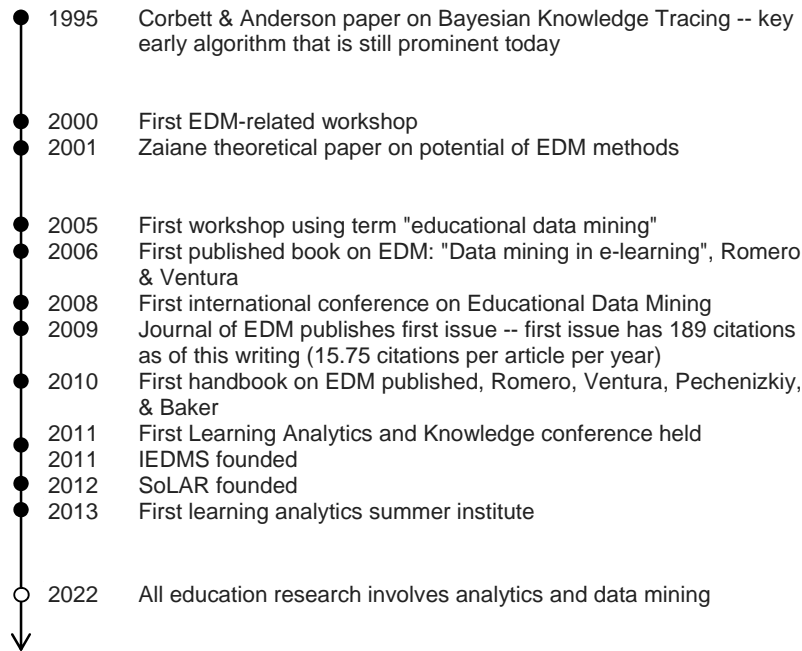email: baker2@exchange.tc.columbia.edu

●	1995	Corbett & Anderson paper on Bayesian Knowledge Tracing -- key early algorithm that is still prominent today
●	2000	First EDM-related workshop
●	2001	Zaiane theoretical paper on potential of EDM methods
●	2005	First workshop using term "educational data mining"
●	2006	First published book on EDM: "Data mining in e-learning", Romero & Ventura
●	2008	First international conference on Educational Data Mining
●	2009	Journal of EDM publishes first issue -- first issue has 189 citations as of this writing (15.75 citations per article per year)
●	2010	First handbook on EDM published, Romero, Ventura, Pechenizkiy, & Baker
●	2011	First Learning Analytics and Knowledge conference held
●	2011	IEDMS founded
●	2012	SoLAR founded
●	2013	First learning analytics summer institute
○	2022	All education research involves analytics and data mining

Fig. X.1. Timeline of significant milestones in EDM

differences given in (Siemens and Baker, 2012). In that work, it was argued that there are five key areas of difference between the communities, including a preference for automated paradigms of data analysis (EDM) versus making human judgment central (LA), a reductionist focus (EDM) versus a holistic focus (LA), and a comparatively greater focus on automated adaptation (EDM) versus supporting human intervention (LA). Siemens and Baker noted that these differences reflected general trends in the two communities rather than hard-and-fast rules. They also noted differences in preferred methodology between the two communities, a topic which we will return to throughout this chapter. Another perspective on the difference between the communities was offered in a recent talk by John Behrens at the LAK 2012 conference, where Dr. Behrens stated that (somewhat contrary to the names of the two communities), EDM has a greater focus on learning as a research topic, while learning analytics has a greater focus on aspects of education beyond learning. In our view, the overlap and differences between the communities is largely organic, developing from the interests and values of specific researchers rather than reflecting a deeper philosophical split or antagonism.

In this remainder chapter, we will review the key methods of educational data mining and ongoing trends, returning to the issue of how EDM compares methodologically to learning analytics as we do so.

X.2 Key EDM Methods

A wide range of EDM methods have emerged through the last several years. Some are roughly similar to those seen in the use of data mining in other domains, whereas others are unique to educational data mining. In this section we will discuss four major classes of method that are in particularly frequent use by the EDM community, including: (a) Prediction Models, (b) Structure Discovery, (c) Relationship Mining, and (d) Discovery with Models. This is not an exhaustive selection of EDM methods; more comprehensive reviews can be found in (Baker and Yacef, 2009; Romero and Ventura, 2007, 2010; Scheuer and McLaren, 2011). Instead, we focus on a subset of methods that are in particularly wide use within the EDM community.

X2.1 Prediction methods

In prediction, the goal is to develop a model which can infer a single aspect of the data (the predicted variable, similar to dependent variables in traditional statistical analysis) from some combination of other aspects of the data (predictor variables, similar to independent variables in traditional statistical analysis).

In EDM, classifiers and regressors are the most common types of prediction models, and each has several subtypes, which we will discuss below. Classifiers and regressors have a rich history in data mining and artificial intelligence, which is leveraged by EDM research. The area of latent knowledge estimation is of particular importance within EDM, and work in this area largely emerges from the User Modeling, Artificial Intelligence in Education, and Psychometrics/Educational Measurement traditions.

Prediction requires having labels for the output variable for a limited data set, where a label represents some trusted ground truth information about the predicted variable's value in specific cases. Ground truth can come from a variety of sources, including "natural" sources such as whether a student chooses to drop out of college (Dekker et al., 2009), state standardized exam scores (Feng, Heffernan and Koedinger, 2009), or grades assigned by instructors, and in approaches where labels are created solely to use as ground truth, using methods such as self-report (cf. D'Mello et al., 2008), video coding (cf. D'Mello et al., 2008), field observations (Baker, Corbett and Koedinger, 2004), and text replays (Sao Pedro et al., 2010).

Prediction models are used for several applications. They are most commonly used to predict what a value will be in contexts where it is not desirable to directly obtain a label for that construct. This is particularly useful if it can be conducted in real time, for instance to predict a student's knowledge (cf. Corbett and Anderson, 1995) or affect (D'Mello et al., 2008; Baker et al., 2012) to support intervention, or to predict a student's future outcomes (Dekker et al., 2009; San Pedro et al., in

press). Prediction models can also be used to study which specific constructs play an important role in predicting another construct (for instance, which behaviors are associated with the eventual choice to attend high school) (cf. San Pedro et al., in press).

X.2.1.1 Classification

In classifiers, the predicted variable can be either a binary or categorical variable. Some popular classification methods in educational domains include decision trees, random forest, decision rules, step regression, and logistic regression. Note that step regression and logistic regression, despite their names, are classifiers rather than regressors. In EDM, classifiers are typically validated using cross-validation, where part of the data set is repeatedly and systematically held out and used to test the goodness of the model. Cross-validation should be conducted at multiple levels, in line with what type of generalizability is desired; for instance, it is typically standard in EDM for researchers to cross-validate at the student level in order to ensure that the model will work for new students, although researchers also cross-validate in terms of populations or learning content. Some common metrics used for classifiers include A'/AUC (Hanley and McNeil, 1982), kappa (Cohen, 1960), precision (Davis and Goadrich, 2006), and recall (Davis and Goadrich, 2006); accuracy, often popular in other fields, is not sensitive to base rates and should only be used if base rates are also reported.

X.2.1.2 Regression

In regression, the predicted variable is a continuous variable. The most popular regressor within educational data mining is linear regression (note that a model produced through this method is mathematically the same as linear regression as used in statistical significance testing, but that the method for selecting and validating the model in EDM use of linear regression is quite different than in statistical significance testing), with regression trees also fairly popular. Regressors such as neural networks and support vector machines, which are prominent in other data mining domains, are somewhat less common in educational data mining. This is thought to be because the high degrees of noise and multiple explanatory factors in educational domains often lead to more conservative algorithms being more successful. Regressors can be validated using the same overall techniques as that in classifiers, often using the metrics of linear correlation or root mean squared error (RMSE).

X.2.1.3 Latent Knowledge Estimation

One special case of classification that is particularly important in educational data mining is latent knowledge estimation. In latent knowledge estimation, a student's knowledge of specific skills and concepts is assessed by their patterns of correctness on those skills (and occasionally other information as well). The word "latent" refers to the idea that knowledge is not directly measurable, it must be inferred from a student's performance. Inferring a student's knowledge can be useful for many goals – it can be a meaningful input to other analyses (we discuss this use below, in the section on discovery with models), it can be useful for deciding when to advance a student in a curriculum (Corbett & Anderson, 1995) or intervene in other ways (cf. Roll et al., 2007), and it can be very useful information for instructors (Feng & Heffernan, 2007).

The models used for estimating latent knowledge in online learning typically differ from the psychometric models used in paper tests or in computer-adaptive testing, as the latent knowledge in online learning is itself dynamic. The models used for latent knowledge estimation in EDM come from two sources: new takes on classical psychometric approaches, and the user modeling/artificial intelligence in education literature. A wide range of algorithms exist for latent knowledge estimation. The classic algorithm is either Bayes Nets (Martin and VanLehn, 1995; Shute, 1995) for complex knowledge structures, or Bayesian Knowledge Tracing (Corbett and Anderson, 1995) for cases where each problem or problem step is primarily associated with a single skill at the point in time when it is encountered. Recently, there has also been work suggesting that an approach based on logistic regression, Performance Factors Assessment (Pavlik et al., 2009) can be effective for cases where multiple skills are relevant to a problem or problem step at the same time. Work by Pardos and colleagues (2012) has also found evidence that combining multiple approaches through ensemble selection can be more effective for large data sets than single models.

X.2.2 Structure Discovery

Structure discovery algorithms attempt to find structure in the data without any ground truth or a priori idea of what should be found. In this way, this type of data mining contrasts strongly with prediction models, above, where ground truth labels must be applied to a subset of the data before model development can occur. Common structure discovery algorithms in educational data include clustering, factor analysis, and domain structure discovery algorithms. Clustering and factor analysis have been used since the early days of the field of statistics, and were refined and explored further by the data mining and machine learning communities. Domain structure discovery emerged from the field of psychometrics/educational measurement.

A fourth type of structure discovery, Network Analysis, is more characteristic of work in learning analytics than in educational data mining (cf. Dawson, 2008; Suthers and Rosen, 2011), and is not discussed in detail here for that reason.

As methods that discover structure without ground truth, less attention is generally given to validation than in prediction, though goodness and fit calculations are still used in determining if a specific structure is superior to another structure.

X.2.2.1 Clustering

In clustering, the goal is to find data points that naturally group together, splitting the full data set into a set of clusters (Kaufman and Rousseeuw, 1990). 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, each data point in a cluster 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 grain-sizes. For example, schools could be clustered together (to investigate similarities and differences among schools), students could be clustered together (to investigate similarities and differences among students), or student actions could be clustered together (to investigate patterns of behavior) (cf. Amershi and Conati, 2009; Beal, Qu and Lee, 2006). Clustering algorithms typically split into two categories: hierarchical approaches such as hierarchical agglomerative clustering (HAC), and non-hierarchical approaches such as k-means, gaussian mixture modeling (sometimes referred to as EM-based clustering), and spectral clustering. The key difference is that hierarchical approaches assume that clusters themselves cluster together, whereas non-hierarchical approaches assume that clusters are separate from each other.

X.2.2.2 Factor Analysis

In factor analysis, the goal is to find variables that naturally group together, splitting the set of variables (as opposed to the data points) into a set of latent (not directly observable) factors (Kline, 1993). Factor analysis is frequently used in psychometrics for validating or determining scales. In educational data mining, factor analysis is used for dimensionality reduction (e.g., reducing the number of variables), including in pre-processing to reduce the potential for over-fitting and to determine meta-features. One example of its use in educational data mining is work to determine which features of intelligent tutoring systems group together (cf. Baker et al., 2009); another example is as a step in the process of developing a prediction model (cf. Minaei-Bidgoli, Kashy, Kortemeyer and Punch, 2003). Factor analysis includes algorithms such as principal component analysis and exponential-family principal components analysis.

X.2.2.3 Domain Structure Discovery

Domain structure discovery consists of finding which items map to specific skills across students. The Q-Matrix approach for doing so is well-known in psychometrics (cf. Tatsuoka, 1995). Considerable work has recently been applied to this problem in educational data mining, for both test data (cf. Barnes, Bitzer and Vouk, 2005; Desmarais, 2011), and for data tracking learning during use of an intelligent tutoring system (Cen, Koedinger and Junker, 2006). Domain structures can be compared using information criteria metrics (Koedinger, McLaughlin and Stamper, 2011), which assess fit compared to the complexity of the model (more complex models should be expected to spuriously fit data better). A range of algorithms can be used for domain structure discovery, from purely automated algorithms (cf. Barnes, Bitzer and Vouk, 2005; Desmarais, 2011; Thai-Nghe, Horváth and Schmidt-Thieme, 2011), to approaches that utilize human judgment within the model discovery process such as Learning Factors Analysis (LFA; Cen et al., 2006).

X.2.3 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 in common use in EDM: association rule mining, sequential pattern mining, correlation mining, and causal data mining. Association rule mining comes from the field of data mining, in particular from "market basket" analysis used in mining of business data (Brin et al., 1997); sequential pattern mining also comes from data mining, with some variants emerging from the bioinformatics community; correlation mining has been a practice in statistics for some time (and the methods of post-hoc analysis came about in part to make this type of method more valid); causal data mining also comes from the intersection of statistics and data mining (Spirtes, Glymour and Scheines, 2000).

X.2.3.1 Association Rule 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:

    IF student is frustrated **OR** has a stronger goal of learning than performance
    THEN the student frequently asks for help

Rules mined using this approach uncover common co-occurrences in data which would have been difficult to discover manually. Association rule mining has been used for a variety of applications in EDM. For example, Ben-Naim and colleagues (2009) found association rules within student data from an engineering data, representing patterns of successful student performance, and Merceron and Yacef (2005) studied which student errors tend to go together.

There is ongoing debate as to which metrics lead to finding the most interesting and usable association rules; a discussion of this issue can be found in Merceron and Yacef (2008), who recommend in particular cosine and lift.

X.2.3.2 Sequential Pattern Mining

In sequential pattern mining, the goal is to find temporal associations between events. Two paradigms are seen that find sequential patterns -- classical sequential pattern mining (Srikant and Agrawal, 1996), which is a special case of association rule mining, and motif analysis (Lin et al., 2002), a method often used in bioinformatics to find common general patterns that can vary somewhat. These methods, like association rule mining, have been used for a variety of applications, including to study what paths in student collaboration behaviors leads to a more successful eventual group project (Perera et al., 2009), the patterns in help-seeking behavior over time (Shanabrook et al., 2010), and studying which patterns in the use of concept maps are associated with better overall learning (Kinnebrew and Biswas, 2012). Sequential pattern mining algorithms, like association rule mining algorithms, depend on a number of parameters to select which rules are worth outputting.

X.2.3.3 Correlation Mining

In correlation mining, the goal is to find positive or negative linear correlations between variables. This goal is not a new one; it is a well-known goal within statistics, where a literature has emerged on how to use post-hoc analysis and/or dimensionality reduction techniques in order to avoid finding spurious relationships. The False Discovery Rate paradigm (cf. Benjamini and Hochberg, 1995; Storey, 2003) has become increasingly popular among data mining researchers, across domains. Correlation mining has been used to study the relationship between student attitudes and help-seeking behaviors (Arroyo and Woolf, 2005; Baker, Walonoski, et al., 2008), and to study the relationship between the design of intelligent tutoring systems and whether students game the system (Baker et al., 2009).

X.2.3.4 Causal Data Mining

In causal data mining, the goal is to find whether one event (or observed construct) was the cause of another event (or observed construct) (Spirtes et al., 2001). Causal data mining is distinguished from prediction in its attempts to find not just predictors but actual causal relationships, through looking at the patterns of covariance between those variables and other variables in the data set. Causal data mining in packages such as TETRAD (Scheines et al., 1998) has been used in EDM to predict which factors will lead a student to do poorly in a class (Fancsali, 2012), to analyze how different conditions of a study impact help use and learning differently (Rau and Scheines, 2012), and to study how gender and attitudes impact behaviors in an intelligent tutor and consequent learning (Rai and Beck, 2011).

X.2.4 Discovery with Models

In discovery with models, 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 a second analysis or model, for example in prediction or relationship mining. Discovery with models is not common in data mining in general, but is seen in some form in many computational science domains.

In the case of educational data mining, one common use is when an initial model's predictions (which represent predicted variables in the original model) become predictor variables in a new prediction model. For instance, prediction models of student robust learning have generally depended on models of student meta-cognitive behaviors (cf. Baker, Gowda and Corbett, 2011a, 2011b), which have in turn depended on assessments of latent student knowledge (cf. Aleven et al., 2006). These assessments of student knowledge have in turn depended on models of domain structure.

When using relationship mining, the relationships between the initial model's predictions and additional variables are studied. This enables a researcher to study the relationship between a complex latent construct and a wide variety of observable constructs, for example investigating the relationship between gaming the system (as detected by an automated detector) and student individual differences (Baker, Walonoski, et al., 2008).

Often, discovery with models leverages the generalization of a prediction model across contexts. For instance, Baker and Gowda (2010) used predictions of gaming the system, off-task behavior, and carelessness across a full year of educational software data to study the differences in these behaviors between an urban, rural, and suburban school in the same region.

X.3 Trends in EDM methodologies and research

Given that "educational data mining" has been around as a term for almost a decade at this writing, and several early EDM researchers had been working in this area even before the community had begun to coalesce, we can begin to see trends and changes in emphasis occurring over time.

One big shift in EDM is the relative emphasis given to relationship mining. In the early years of EDM, relationship mining was used in almost half of the articles published (Baker and Yacef, 2009). Relationship mining methods have continued to be important in EDM since then, but it is fair to say that the dominance of relationship mining has reduced somewhat in the following years. For example in the EDM2012 conference, only 16% of papers use relationship mining as defined in this article.

Prediction and clustering were important methods in the early years of EDM (Baker and Yacef, 2009), and have continued to be highly used. However, within the category of prediction modeling, the distribution of methods has changed substantially. Classification and regression were important in 2005-2009, and remain important to this day, but latent knowledge estimation has increased substantially in importance, with articles representing different paradigms for how to estimate student knowledge competing to see which algorithms are most effective in which contexts (Pavlik et al., 2009; Gong, Beck and Heffernan, 2011; Pardos et al., 2012).

A related trend is the increase in the prominence of domain structure discovery in recent EDM research. Although domain structure discovery has been part of EDM from the beginning (Barnes, 2005), recent years have seen increasing work on a range of approaches for modeling domains. Some work has attempted to find better ways to find q-matrices expressing domain structure in a purely empirical fashion (Desmarais, 2011; Desmarais et al, 2012), while other work attempts to leverage human judgment in fitting q-matrices (Cen, Koedinger and Junker, 2007; Koedinger, McLaughlin and Stamper, 2012). Additionally, in recent years there has been work attempting to automatically infer prerequisite structures in data (Beheshti and Desmarais, 2012), and to study the impact of not following prerequisite structures (Vuong et al., 2011).

A third emerging emphasis in EDM is the continued trend towards modeling a greater range of constructs. Though the trends in latent knowledge estimation and domain structure discovery reflect the continued emphasis within EDM on modeling student knowledge and skill, there has been a simultaneous trend towards expanding the space of constructs modeled through EDM, with researchers expanding from modeling knowledge and learning to modeling constructs such as metacognition, self-regulation, motivation, and affect (cf. Goldin et al., 2012; Bouchet et al., 2012; Baker et al., 2012). The increase in the range of constructs being modeled in EDM has been accompanied by an increase in the number of discovery with models analyses leveraging those models to support basic discovery.

X.4 EDM and Learning Analytics

Many of the same methodologies are seen in both EDM and Learning Analytics. Learning analytics has a relatively greater focus on human interpretation of data and visualization (though there is a tradition of this in EDM as well -- cf. Kay et al., 2006; Martinez et al., 2011). EDM has a relatively greater focus on automated methods. But ultimately, in our view, the differences between the two communities are more based on focus, research questions, and the eventual use of models (cf. Siemens and Baker, 2012), than on methods used.

Prediction models are prominent in both communities, for instance, although Learning Analytics researchers tend to focus on classical approaches of classification and regression more than on latent knowledge estimation. Structure Discovery is prominent in both communities, and in particular clustering has an important role in both communities. In terms of specialized/domain-specific structure discovery algorithms, domain structure discovery is more emphasized by EDM researchers while network analysis/social network analysis is more emphasized in learning analytics (Bakharia and Dawson, 2011; Schreurs et al., 2013), again more due to research questions adopted by specific researchers, than a deep difference between the fields. Relationship mining methods are significantly more common in EDM than in learning analytics. It is not immediately clear to the authors of this paper why relationship mining methods have been less utilized in learning analytics than in EDM, given the usefulness of these methods for supporting interpretation by analysts (this point is made in d'Aquin and Jay, 2013, who demonstrate the use of sequential pattern mining in learning analytics). Discovery with models is significantly more common in EDM than learning analytics, and much of its appearance at LAK conferences is in papers written by authors more known as members of the EDM community (e.g. Pardos et al., 2013). This is likely to again be due to differences in research question and focus; even though both communities use prediction modeling, LAK papers tend to predict larger constructs (such as dropping out and course failure) whereas EDM papers tend to predict smaller constructs (such as boredom and short-term learning), which are more amenable to then use in discovery with models analyses of larger constructs.

Finally, some methodological areas are more common in learning analytics than in educational data mining (though relatively fewer, owing to the longer history of EDM). The most prominent example is the automated analysis of textual data. Text analysis, text mining, and discourse analysis is a leading area in learning analytics; it is only seen occasionally in EDM (cf. D'Mello, Olney and Person, 2010; Rus et al., 2012).

X.5 Conclusion

In recent years, two communities have grown around the idea of using large-scale educational data to transform practice in education and education research. As this area emerges from relatively small and unknown conferences to a theme that is known throughout education research, and which impacts schools worldwide, there is an opportunity to leverage the methods listed above to accomplish a variety of goals. Every year, the potential applications of these methods become better known, as researchers and practitioners utilize these methods to study new constructs and answer new research questions.

While we learn where these methods can be applied, we are also learning how to apply them more effectively. Having multiple communities and venues to discuss these issues is beneficial; having communities that select work with different values and perspectives will support the development of a field that most effectively uses large-scale educational data. Ultimately, the question is not which methods are best, but which methods are useful for which applications, in order to improve the support for any person who is learning, whenever they are learning.

X.6 References

Aleven V, Mclaren B, Roll I, Koedinger K (2006) Toward meta-cognitive tutoring: A model of help seeking with a cognitive tutor. International journal of artificial intelligence in education 16(2):101–128.

Amershi S, Conati C (2009) Combining unsupervised and supervised classification to build user models for exploratory learning environments. Journal of educational data mining 1(1):18–71.

Arroyo I, Woolf B (2005) Inferring learning and attitudes from a bayesian network of log file data. In: Proceedings of the 12th international conference on artificial intelligence in education, pp 33–40.

Baker R, Gowda S (2010) An analysis of the differences in the frequency of students' disengagement in urban, rural, and suburban high schools. In: proceedings of the 3rd international conference on educational data mining, pp 11–20.

Baker R, Yacef K (2009) The state of educational data mining in 2009: A review and future visions. Journal of educational data mining 1(1):3–17.

Baker R, Corbett AT, Koedinger K, Wagner AZ (2004) Off-task behavior in the cognitive tutor classroom: when students game the system. In: Proceedings of the SIGCHI conference on human factors in computing systems, pp 383–390.

Baker R, Walonoski J, Heffernan N, Roll I, Corbett A, Koedinger K (2008) Why students engage in " gaming the system" behavior in interactive learning environments. Journal of interactive learning research 19(2):185–224.

Baker R, de Carvalho A, Raspat J, Aleven V, Corbett A, Koedinger K (2009) Educational software features that encourage and discourage "gaming the system". In: Proceedings of the international conference on artificial intelligence in education, pp 475–482.

Baker R, Gowda SM, and Corbett AT (2011a). Towards predicting future transfer of learning. In: Artificial Intelligence in education. Lecture notes in computer science, vol 6738. pp 23-30

Baker R R, Gowda SM, and Corbett AT (2011b). Automatically detecting a student's preparation for future learning: Help use is key. In: Proceedings of the 4th international conference on educational data mining, pp 179–188.

Baker R, Kalka J, Aleven V, Rossi L, Gowda S, Wagner A, Kusbit G, Wixon M, Salvi A, Ocumpaugh J (2012) Towards sensor-free affect detection in cognitive tutor algebra. In: Proceedings of the 5th international conference on educational data mining, pp 126–133.

Bakharia A, Dawson S (2011) SNAPP: a bird's-eye view of temporal participant interaction. In: Proceedings of the 1st international conference on learning analytics and knowledge, pp 168–173.

Barnes T (2005) The q-matrix method: Mining student response data for knowledge. In: Proceedings of the American association for artificial intelligence 2005 educational data mining workshop, pp 39–46.

Barnes T, Bitzer D, Vouk M (2005) Experimental analysis of the q-matrix method in knowledge discovery. In: Foundations of intelligent systems. Lecture notes in computer science, vol 3488. pp 603-611.

Beal CR, Qu L, Lee H (2006) Classifying learner engagement through integration of multiple data sources. In: Proceedings of the 21st national conference on artificial intelligence, pp 151–156.

Beheshti B, Desmarais M (2012) Improving matrix factorization techniques of student test data with partial order constraints. In: Proceedings of the 20th international conference on user modeling, adaptation, and personalization, pp 346–350.

Ben-Naim D, Bain M, Marcus N (2009) A user-driven and data-driven approach for supporting teachers in reflection and adaptation of adaptive tutorials. In: Proceedings of the 2nd International Conference on Educational Data Mining, pp 21-30.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the royal statistical society 57(1): 289–300.

Bouchet F, Azevedo R, Kinnebrew J, Biswas G (2012) Identifying students' characteristic learning behaviors in an intelligent tutoring system fostering Self-Regulated learning. In: Proceedings of the 5th international conference on educational data mining, pp 65–72.

Brin S, Motwani R, Ullman J, Tsur S (1997) Dynamic itemset counting and implication rules for market basket data. In: Proceedings of the 1997 ACM international conference on management of data, pp 255–264.

Cen H, Koedinger K, Junker B (2006) Learning factors analysis–a general method for cognitive model evaluation and improvement. In: Intelligent tutoring systems. Lecture notes in computer science, vol 4053. pp 164–175.

Cen H, Koedinger K, Junker B (2007) Is over practice necessary?-improving learning efficiency with the cognitive tutor through educational data mining. In: Proceedings of 13th international conference on artificial intelligence in education, pp 511–518.

Cen H, Koedinger K, Junker B (2008) Comparing two IRT models for conjunctive skills. In: Proceedings of the 9th international conference on intelligent tutoring systems, pp 796–798.

Cohen J (1960) A coefficient of agreement for nominal scales. Educational and psychological measurement 20(1):37–46.

Corbett A, Anderson J (1995) Knowledge tracing: Modeling the acquisition of procedural knowledge. User modeling and user-adapted interaction 4(4):253–278.

Craig S, D'Mello S, Witherspoon A, Graesser A (2008) Emote aloud during learning with AutoTutor: Applying the facial action coding system to cognitive–affective states during learning. Cognition and emotion 22(5):777–788.

d'Aquin M, Jay N (2013) Interpreting data mining results with linked data for learning analytics: motivation, case study and directions. In: Proceedings of the 3rd international conference on learning analytics and knowledge, pp 155–164.

Davis J, Goadrich M (2006) The relationship between Precision-Recall and ROC curves. In: Proceedings of the 23rd international conference on machine learning, pp 233–240.

Dawson S (2008) A study of the relationship between student social networks and sense of community. Educational technology and society 11(3):224–238.

Dekker, G, Pechenizkiy M, Vleeshouwers J (2009) Predicting students drop out: a case study. In: Proceedings of 2nd international conference on educational data mining, pp 41–50.

Desmarais M (2011) Conditions for effectively deriving a q-matrix from data with non-negative matrix factorization. In: Proceedings of the 4th international conference on educational data mining, pp 41–50.

Desmarais M, Beheshti B, Naceur R (2012) Item to skills mapping: deriving a conjunctive q-matrix from data. In: Intelligent Tutoring Systems. Lecture notes in computer science, vol 7315. pp 454–463.

D'Mello S, Craig S, Witherspoon A, Mcdaniel B, Graesser A (2008) Automatic detection of learner's affect from conversational cues. User modeling and user-adapted interaction 18(1-2):45–80.

D'Mello S, Olney A, Person N (2010) Mining collaborative patterns in tutorial dialogues. Journal of educational data mining 2(1):1–37.

Fancsali S (2012) Variable construction and causal discovery for cognitive tutor log data: Initial results. In: Proceedings of the 5th conference on educational data mining, pp 238–239.

Feng M, Heffernan N (2007) Towards Live Informing and Automatic Analyzing of Student Learning: Reporting in the Assistment System. Journal of Interactive Learning Research 18 (2): 207-230.

Feng M, Heffernan N, Koedinger K (2009) Addressing the assessment challenge with an online system that tutors as it assesses. User modeling and user-adapted interaction, 19(3):243–266.

Goldin I, Koedinger KR, Aleven V (2012) Learner differences in hint processing. In: Proceedings of the 5th international conference on educational data mining, pp 73–80.

Gong Y, Beck JE, Heffernan NT (2011) How to construct more accurate student models: comparing and optimizing knowledge tracing and performance factor analysis. International journal of artificial intelligence in education 21(1):27–46.

Hanley A, McNeil B (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143(1):29–36.

Kaufman L, Rousseeuw PJ(1990) Finding Groups in Data: An Introduction to Cluster Analysis.

Kay J, Maisonneuve N, Yacef K, Zaïane O (2006) Mining patterns of events in students' teamwork data. In: Proceedings of the workshop on educational data mining at the 8th international conference on intelligent tutoring systems, pp 45–52.

Kinnebrew J, Biswas G (2012) Identifying learning behaviors by contextualizing differential sequence mining with action features and performance evolution. In: Proceedings of the 5th international conference on educational data mining, pp 57–64.

Kline P (1993) An Easy Guide to Factor Analysis.

Koedinger K, Pavlik P, Stamper J, Nixon T, Ritter S (2010) Avoiding problem selection thrashing with conjunctive knowledge tracing. In: Proceedings of the 3rd international conference on educational data mining, pp 91–100.

Koedinger K, McLaughlin E, Stamper J (2012) Automated student model improvement. In: Proceedings of the 5th international conference on educational data mining, pp 17–24.

Lin J, Keogh E, Lonardi S, Patel P (2002) Finding motifs in time series. In: Proceedings of the 2nd workshop on temporal data mining, pp 53-68.

Martin J, VanLehn K (1995) Student assessment using bayesian nets. International journal of human computer studies 42(6):575–592.

Martinez R, Yacef K, Kay J, Kharrufa A, Al-Qaraghuli A (2011) Analysing frequent sequential patterns of collaborative learning activity around an interactive tabletop. In: Proceedings of the 4th international conference on educational data mining, pp 111-120.

Merceron A, Yacef K (2005) Educational data mining: a case study. In: Proceedings of the 2005 conference on artificial intelligence in education: supporting learning through socially informed technology, pp 467–474.

Merceron A, Yacef K (2008) Interestingness measures for association rules in educational data. In: Proceedings of the 1st international conference on educational data mining, pp 57–66.

Minaei-Bidgoli B, Kashy D, Kortmeyer G, Punch W (2003) Predicting student performance: an application of data mining methods with an educational web-based system. In: Frontiers in education, 2003. FIE 2003 33rd Annual, pp T2A 13–18.

Pardos ZA, Gowda SM, Baker R, Heffernan NT (2012) The sum is greater than the parts: ensembling models of student knowledge in educational software. ACM SIGKDD explorations newsletter 13(2):37–44.

Pardos Z, Baker R, San Pedro M, Gowda S, Gowda S (2013) Affective states and state tests: investigating how affect throughout the school year predicts end of year learning outcomes. In: Proceedings of the 3rd international conference on learning analytics and knowledge, pp 117–124.

Pavlik P, Cen H, Koedinger KR (2009) Performance factors analysis - a new alternative to knowledge tracing. In: Proceedings of the 14th international conference on artificial intelligence in education, pp 531–538.

Perera D, Kay J, Koprinska I, Yacef K, Zaïane OR (2009) Clustering and sequential pattern mining of online collaborative learning data. Knowledge and data engineering, IEEE transactions on 21(6):759–772.

Rai D, Beck J (2011) Exploring user data from a game-like math tutor: A case study in causal modeling. In: Proceedings of the 4th international conference on educational data mining, pp 307-313.

Rau A, Scheines R (2012) Searching for variables and models to investigate mediators of learning from multiple representations. In: Proceedings of the 5th international conference on educational data mining, pp 110-117.

Roll, I., Aleven, V., McLaren, B.M., Koedinger, K.R. (2007) Can help seeking be tutored? Searching for the secret sauce of metacognitive tutoring. In: Proceedings of the 13th International Conference on Artificial Intelligence in Education, pp. 203–210. Marina del Rey, CA.

Romero C, Ventura S (2007) Educational data mining: A survey from 1995 to 2005. Expert Systems with Applications 33(1):135–146.

Romero C, Ventura S (2010) Educational data mining: A review of the state of the art. Systems, Man, and Cybernetics, Part C: Applications and reviews, IEEE transactions on 40(6):601–618.

Rus V, Moldovan C, Graesser A, Niraula N (2012) Automated discovery of speech act categories in educational games. In: Proceedings of the 5th international conference on educational data mining, pp 25–32.

San Pedro M, Baker R, Bowers A, and Heffernan N (2013) Predicting College Enrollment from Student Interaction with an Intelligent Tutoring System in Middle School. To appear In: Proceedings of the 6th international conference on educational data mining (in press)

Sao Pedro M, Baker R, Montalvo O, Nakama A, Gobert JD (2010) Using text replay tagging to produce detectors of systematic experimentation behavior patterns. In: Proceedings of the 3rd international conference on educational data mining, pp 181–190.

Scheines R, Spirtes P, Glymour C, Meek C, Richardson T (1998) The TETRAD project: Constraint based aids to causal model specification. Multivariate Behavioral Research 33(1):65–117.

Scheuer O, McLaren BM (2011) Educational data mining. The encyclopedia of the sciences of learning.

Schreurs B, Teplovs C, Ferguson R, De Laat M, Buckingham Shum S (2013) Visualizing social learning ties by type and topic: rationale and concept demonstrator. In: Proceedings of the 3rd international conference on learning analytics and knowledge, pp 33-37.

Shanabrook DH, Cooper DG, Woolf BP, Arroyo I (2010) Identifying high-level student behavior using sequence-based motif discovery. In: Proceedings of the 3rd international conference on educational data mining, pp 191-200.

Shute VJ (1995) SMART: Student modeling approach for responsive tutoring. User modeling and user-adapted interaction, 5(1):1–44.

Siemens G, Baker R (2012) Learning analytics and educational data mining: towards communication and collaboration. In: Proceedings of the 2nd international conference on learning analytics and knowledge, pp 252–254.

Spirtes P, Glymour C, Scheines R (2000) Causation, Prediction, and Search.

Srikant R, Agrawal R (1996) Mining sequential patterns: Generalizations and performance improvements. Research report: IBM Research Division.

Storey JD (2003) The positive false discovery rate: A bayesian interpretation and the q-value. Annals of statistics 31(6): 2013–2035.

Suthers D, Rosen D (2011) A unified framework for multi-level analysis of distributed learning. In: Proceedings of the 1st international conference on learning analytics and knowledge. 64-74.

Thai-Nghe N, Horvath T, Schmidt-Thieme L (2011) Context-Aware factorization for personalized student's task recommendation. In: Proceedings of the international workshop on personalization approaches in learning environments, pp 13-18.

Tatsuoka K (1995) Architecture of knowledge structures and cognitive diagnosis: A statistical pattern recognition and classification approach. In: Cognitively diagnostic assessment, pp 327–359.

Vuong A, Nixon T, Towle B (2011) A method for finding prerequisites within a curriculum. In: Proceedings of the 4th international conference on educational data mining, pp 211-216.