Mining Data for Student Models

Ryan S.J.d. Baker[1]

[1] Department of Social Science and Policy Studies, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609 USA
rsbaker@wpi.edu

Abstract. Data mining methods have in recent years enabled the development of more sophisticated student models which represent and detect a broader range of student behaviors than was previously possible. This chapter summarizes key data mining methods that have supported student modeling efforts, discussing also the specific constructs that have been modeled with the use of educational data mining. We also discuss the relative advantages of educational data mining compared to knowledge engineering, and key upcoming directions that are needed for educational data mining research to reach its full potential.

Keywords: Educational Data Mining, Data Mining, Text Replays, Knowledge Engineering

1 Introduction

In recent years, student models within intelligent tutoring systems have expanded to include an impressive breadth of constructs about the individual student (or pairs or groups of students), with increasing levels of precision. Student models increasingly assess not just whether a student knows a skill or concept, but a broad range of affective, meta-cognitive, motivational, and behavioral constructs. These advances in student modeling, discussed in some detail in earlier chapters, have been facilitated by advances in educational data mining methods [15, 50] that leverage fine-grained data about student behavior and performance. As this data becomes increasingly available at large-scale [cf. 18, 39], there are increasing opportunities for developing increasingly sophisticated and broad-based models of the student using an intelligent tutoring system.

In this chapter, we give a brief overview of educational data mining methodologies, focusing on techniques that have seen specific application in order to enrich and otherwise improve student models. We also discuss the key differences between educational data mining and knowledge engineering approaches to enriching student models, and discuss key steps that would facilitate the more rapid application of educational data mining methods to a broader range of educational software and learner constructs.

Data mining, also called Knowledge Discovery in Databases (KDD), is the field of discovering novel and potentially useful information from large amounts of data [59]. On the Educational Data Mining community website, www.educationaldatamining.org, educational data mining (abbreviated as EDM) is defined as: "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."

There are multiple taxonomies of the areas of educational data mining, one by Romero and Ventura [cf. 50], and one by Baker [6], also discussed in [15]. The two taxonomies are quite different, and a full discussion of the differences is out of the scope of this chapter (there is a brief discussion of some of the differences between the taxonomies in [15]). Within this chapter, we focus on the taxonomy from [6, 15], which categorizes research into educational data mining into the following areas:

- Prediction
    - Classification
    - Regression
    - Density estimation
- Clustering
- Relationship mining
    - Association rule mining
    - Correlation mining
    - Sequential pattern mining
    - Causal data mining
- Distillation of data for human judgment
- Discovery with models

Beyond these categories, there is also a continual presence of knowledge engineering methods [32, 54] within educational data mining conferences. Knowledge engineering methods utilize human judgment to create a model of a construct of interest, rather than doing so in an automated fashion. We will discuss the similarities and differences between knowledge engineering and data mining, and the relative advantages and disadvantages of each class of method for developing student models, later in this document.

In this chapter, we emphasize the student modeling applications of prediction methods, clustering methods, and methods for the distillation of data for human judgment. Relationship mining is a key class of method within educational data mining research and is a frequently-used method in EDM venues, as summarized in [15], but is typically focused more on data mining to promote discovery for scientific researchers or other end users, rather than on improving student models. One exception to this is the use of relationship mining to improve models of domain structure [e.g. 46]; recent research along these lines is discussed in detail in another chapter within this volume. Discovery with models research is generally focused on promoting scientific discovery. Discovery with models research involves student models, but in a different direction from the work discussed in this chapter. When discovery with models research involves student models, it leverages the outputs of student models (among other sources) to promote scientific discovery, rather than using automated discovery to promote the improvement of the student models themselves.

Another key way that data mining has influenced student models is by influencing the internal structure of Bayes Nets and Bayesian Knowledge Tracing; these issues are discussed in another chapter within this volume, and are therefore not discussed in detail in this chapter.

2 Prediction Methods

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. Ground truth labels do not need to be perfect in order to be useful for the development of reliable models through data mining; a data mining approach that is not over-fit can accommodate a moderate degree of noise in the original labels, so long as the labels are not systematically biased. Labels which have noise are sometimes referred to as "bronze-standard" labels. The degree of noise in the original labels can often be assessed by assessing the inter-rater reliability of the labels, frequently with Cohen's Kappa [24].

Broadly, there are three types of prediction: classification, regression, and density estimation. Classification and regression historically have played more prominent roles in educational data mining than density estimation. In classification, the predicted variable is a binary or categorical variable. In regression, the predicted variable is quantitative. In both cases, any type of input data is possible, although some algorithms are not able to handle all types of data.

The range of prediction methods used in educational data mining approximately corresponds to the types of prediction methods used in data mining more broadly; however, the techniques emphasized in educational data mining have varied from those most popular in other domains. In particular, support vector machines [53] and neural networks [33], popular methods in other domains, have been relatively less emphasized in educational data mining. Contrastingly, linear methods have been relatively more emphasized. It is not necessarily the case that educational data is particularly likely to be linear – in fact, many have argued that educational and learning data frequently have a non-linear character [40, 56]. However, the relatively high noise in educational data, combined with the relative expense of labeling data in many cases, may bias in favor of approaches which are less likely to over-fit. Over-fitting is when a model does well on the original training data, as the expense of doing poorly on new data [35], in this case, data from new students or new intelligent tutor lessons.

Another distinctive feature of educational data mining research is the use of methods involving modeling frameworks drawn from the psychometrics literature [cf. 41], in combination with machine-learning space-searching techniques [cf. 61] (many examples of this also exist in data mining for domain models, discussed in another chapter in this volume). These methods have the benefit of explicitly accounting for meaningful hierarchy and non-independence in data. For instance, it can be important to detect both which students engage in a behavior of interest, and exactly when (or at least in which parts of the interface) an individual student is engaging in that behavior [cf. 11].

One classification method with prominence both in educational data mining and in data mining within other domains is decision trees [47], used in a considerable amount of educational data mining research [cf. 13, 43, 57]. Decision trees are able to handle both quantitative and categorical features in making model assessment, a benefit given the highly heterogenous feature data often generated by tutors [cf.10, 16, 21, 30, 57], and can explicitly control for over-fitting with a “pruning” step [47].

Educational data mining methods have enabled the construction of student models (or student models components) of a wide number of constructs. For instance, classification and regression methods have been used to develop detectors of gaming the system [11, 13, 57]. These detectors have accurately predicted differences in student learning [23], and have been embedded into intelligent tutoring systems and used to drive adaptive behavior [cf. 10, 58]. Similarly, classification methods have been used to develop detectors of student affect, including frustration, boredom, anxiety, engaged concentration, joy, and distress [25, 30]. Detectors of affect and emotion have been used to drive automated adaptation to differences in student affect, significantly reducing students’ frustration and anxiety [60] and increasing the incidence of positive emotion [22]. Classification methods have also been used to develop detectors of off-task behavior [7, 21], predicting differences in student learning. Additionally, classification methods have been used to infer low self-efficacy [42] and slipping [8].

Classification methods have also enabled improvements to Bayesian Modeling of student knowledge, discussed in another chapter in this volume. For instance, [8] integrates models of student slipping into Bayesian Knowledge Tracing, leading to more accurate prediction of future student performance within Cognitive Tutors.

3 Clustering

In clustering, the goal is to find data points that naturally group together, splitting the full data set into a set of groups, called clusters. Clustering does not require (and does not use) labels of any output variable; the data is clustered based solely on internal similarity, not on any metric of specific interest. 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. The range of clustering methods used in educational data mining approximately corresponds to the types of prediction methods used in data mining more broadly, including algorithms such as k-means [34] and Expectation Maximization (EM)-Based Clustering [19], and model frameworks such as Gaussian Mixture Models [48].

Clustering has been used to develop student models for several types of educational software, including intelligent tutoring systems. In particular, fine-grained models of student behavior at the action-by-action level are clustered in terms of features of the student actions. For instance, Amershi & Conati used clustering on student behavior within an exploratory learning environment, discovering that certain types of reflective behavior and strategic advancement through the learning task were associated with better learning [4]. In addition, Beal and her colleagues applied clustering to study the categories of behavior within an intelligent tutoring system [16]. Other prominent research has investigated how clustering methods can assist in content recommendation within e-learning [55, 62].

Clustering is generally most useful when relatively little is known about the categories of interest in the data set, such as in types of learning environment not previously studied with educational data mining methods [e.g. 4] or for new types of learner-computer interaction, or where the categories of interest are unstable, as in content recommendation [e.g. 55, 62]. The use of clustering in domains where a considerable amount is already known brings some risk of discovering phenomena that are already known. As work in other areas of EDM goes forward, an increasing amount is known about student behavior across learning environments. One potential future use of clustering, in this situation, would be to use clustering as a second stage in the process of modeling student behavior in a learning system. First, existing detectors could be used to classify known categories of behavior. Then, data points not classified as belonging to any of those known behavior categories could be clustered, in order to search for unknown behaviors. Expectation Maximization (EM)-Based Clustering [19] is likely to be a method of particularly high potential for this, as it can explicitly incorporate already known categories into an initial starting point for clustering.

4 Distillation of Data for Human Judgment

One key recent trend facilitating the use of educational data mining methods to improve student models is the advance in methods for distilling data for human judgment. In many 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 information visualization methods most commonly used within EDM are often different than those most often used for other information visualization problems [cf. 36, 38], owing to the specific structure often present in intelligent tutor data, and the meaning embedded within that structure. For instance, data is meaningfully organized in terms of the structure of the learning material (skills, problems, units, lessons) and the structure of learning settings (students, teachers, collaborative pairs, classes, schools).

Data is distilled for human judgment in educational data mining for two key purposes: classification and identification. One key area of development of data distillations supporting classification is the text replay methodology [12]. An example of a text replay is shown in Figure 1. In this case, sub-sections of a data set are displayed in text format, and labeled by human coders. These labels are then generally used as the basis for the development of a predictor. Text replays are significantly faster than competing methods for labeling, such as quantitative field observations or video coding [12, 13], and achieve good inter-rater reliability [12, 14]. Text replays have been used to support the development of prediction models of gaming the system in multiple learning environments [12, 14], and to develop models of scientific reasoning skill in inquiry learning environments [45, 51]. An alternate approach, displaying a re-constructed replay of a student's screen, has also been used to label student data for use in classification [cf. 29]; however, this approach has become less common, as it is significantly slower than text replays [cf. 12], while not giving more information about student behavior or expression outside the system, unlike methods such as quantitative field observation and video methods.

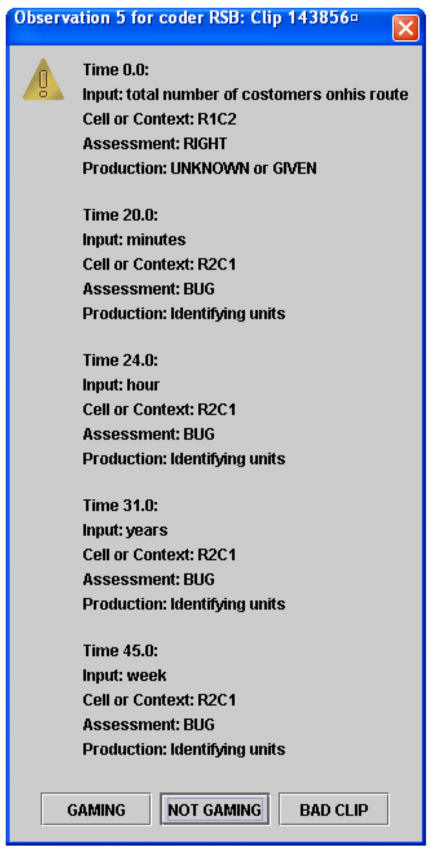

Figure 1. A text replay of student behavior in a Cognitive Tutor (from [13])

Identification of learning patterns and learner individual differences from visualizations is a key method for exploring educational data sets. For instance, Hershkovitz and Nachmias's learnograms provide a rich representation of student behavior over time [36]. Within the domain of student models, a key use of identification with distilled and visualized data is in inference from learning curves, as shown in Figure 2. A great deal can be inferred from learning curves about the character of learning in a domain [26, 39], as well as about the quality of the domain model. Classic learning curves display the number of opportunities to practice a skill on the X axis, and display 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 that successful learning is occurring. A flatter curve, as in Figure 2, indicates that learning is occurring, but with significant difficulty. A sudden spike upwards, by contrast, indicates that more than one knowledge component is included in the model [cf. 26]. A flat high curve indicates poor learning of the skill, and a flat low curve indicates that the skill did not need instruction in the first place. An upwards curve indicates the difficulty is increasing too fast. Hence, learning curves are a powerful tool to support quick inference about the character of learning in an educational system, leading to their recent incorporation into tools used by education researchers outside of the educational data mining community [e.g. 39].

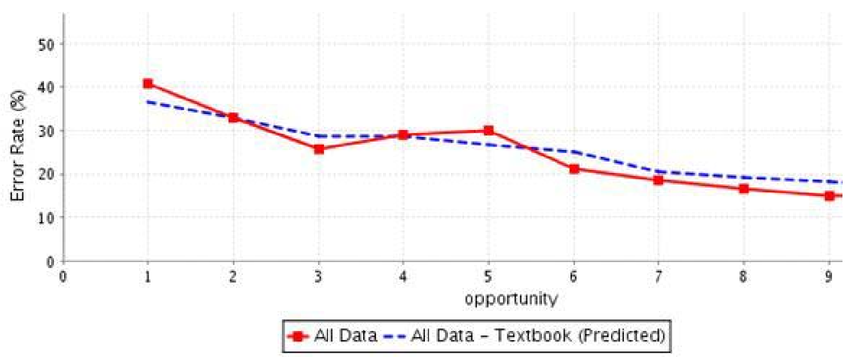

Figure 2. A learning curve of student performance over time in a Cognitive Tutor (from [39])

5 Knowledge Engineering and Data Mining

An alternate method for developing student models is knowledge engineering [32,54]. Knowledge engineering approaches develop models that can engage in problem-solving, reasoning, or decision making, making the same decisions that a human expert would; they can do so simply by replicating the decision-making results or by attempting to develop a cognitive model that reasons in the same fashion that a human expert would. As a method, knowledge engineering relies upon human researchers studying the construct of interest, and directly developing – engineering – the model of the construct of interest. The mapping between features of the data set and the construct of interest is directly made by the engineer. As such, knowledge engineering can be contrasted to classification or regression, which use labels generated through expert decision-making but develop the mapping between the features of the data set and the construct of interest through an automated process.

Knowledge engineering is frequently used to develop domain models, as discussed in another chapter in this volume. Within student modeling, knowledge engineering has been a prominent method for modeling sophisticated student behaviors within intelligent tutoring systems, with a focus on gaming the system and help-seeking behaviors. For instance, Beal and her colleagues used knowledge engineering to model gaming the system [16]. Shih and his colleagues used knowledge engineering to develop a mathematical model that could detect self-explanation and appropriate use of bottom-out hints [52]. Buckley and his colleagues used knowledge engineering to assess students' level of systematicity during problem-solving in interactive simulations [20].

Within student modeling, knowledge engineering is frequently used to develop models of sophisticated student behavior in a hybrid fashion, where knowledge engineering is used to develop the functional form of a mathematical model, and then automated parameter-fitting is used to find (or refine) values for the parameters of that model. For instance, Aleven and his colleagues developed a model of a range of student help-seeking behaviors in Cognitive Tutors [2,3], using knowledge engineering to develop the functional form of a mathematical model, and then automated parameter-fitting to find values for the parameters of that model. Several of the components of Aleven et al's model predicted differences in student learning. In another example, Beck presented a model of hasty guessing [17] (called disengagement in the original paper, but renamed hasty guessing in later work) in an intelligent tutor for reading, developed using knowledge engineering to develop an item-response theory model, and then using automated parameter-fitting to find values for the parameters of that model. Beck's model successfully predicted differences in student post-test scores. Johns & Woolf (2006) used a similar combination of knowledge engineering and parameter fitting to model gaming the system [37].

In addition, educational data mining research often also involves some degree of knowledge engineering during the process of generating the data set features to use within classification or regression. During this step of the data mining process, researchers often attempt infer what types of features an expert coder would use – although this trend is diminishing as features are increasingly re-used in creating new data mining models, either from the same research group, or across research groups [cf. 8, 11, 21, 57].

As can be seen, knowledge engineering and educational data mining have both been used to model gaming the system. Aside from this overlap, the two approaches have been used to model different phenomena, with knowledge engineering methods emphasized in modeling help-seeking while educational data mining methods have been emphasized in modeling affect, self-efficacy, and off-task behavior. It is worth noting that the domains emphasized in educational data mining are often cases where recognition is fairly easy for humans (e.g. it is feasible to tell that a student is bored by looking at him/her [e.g. 31]), but where it is difficult to analyze exactly how those decisions are made in terms of features of data available in the log files. In these cases, an automated process that can test large numbers of alternatives can be considerably more time-efficient than attempting to develop such a mapping through pure rational thought.

In terms of accuracy or effectiveness, educational data mining and knowledge engineering have largely not been directly compared. Mostly they have been used to model different phenomena; even when the same phenomena has been modeled with both educational data mining and knowledge engineering methods, it has been modeled in different learning systems. One exception, however, exists in models of gaming the system within Cognitive Tutors. In [49], Roll and his colleagues compared early versions of knowledge engineered and data mined models of gaming the system [e.g. 2, 9]. Roll and his colleagues found significant correlation between the predictions made by the two models. However, despite that correlation, they found the data mined model was substantially more successful in predicting human labels of gaming behavior than the knowledge engineered model, suggesting that the data mined model had higher construct validity. The comparison used in this paper was not cross-validated, but the data mined model also performed better in predicting student behavior during cross-validation [e.g. 9]. Beyond this, Roll and colleagues found that both models successfully predicted post-test performance, although in both cases the relationship was unstable (see [23] for a discussion of this issue for the data mined model). This suggests that despite the higher construct validity of the data mining model, both models captured the underlying phenomena in qualitatively similar fashions.

This single study, of course, is not conclusive proof that educational data mining achieves higher construct validity than knowledge engineering. In order to investigate this question more thoroughly, it will be necessary to conduct a broader comparison, involving a variety of constructs, and held-out data including new students. One competititon that may produce evidence that can be used to infer the relative accuracy and predictive power of knowledge engineered and data mined methods in educational data will occur in the next year. The 2010 KDD Cup (to be announced at http://www.sigkdd.org/kddcup/index.php) will involve predicting future student correctness in held-out intelligent tutor data from the Pittsburgh Science of Learning Center DataShop [39], and is likely to attract submissions of both types.

6 Key Future Directions

Though educational data mining methods have contributed in significant ways to the sophistication of student models, there are two factors that are currently slowing the extension of these improvements to the full range of intelligent tutoring systems and learner characteristics.

One significant limitation is the relatively low degree of investigation into the generalizability of models between or within intelligent tutoring systems. It is becoming increasingly common to use cross-validation at the student level to verify that a model is applicable to new students; however, it remains rare for researchers to validate that a model generalizes across subsets of an intelligent tutoring curricula. In one example of this type of validation, [11] determined that their data-mined model of gaming the system remained accurate within new intelligent tutor lessons drawn from the same overall system and curricula, using cross-validation at the lesson level. However, few other examples exist in the published literature.

Furthermore, the author of this chapter is not aware of any papers that explicitly study whether any models remain accurate when applied to different tutoring systems. Some papers have studied a related issue, whether the model features utilized in a model of a given construct in one tutoring system are effective in a different tutoring system [e.g. 11, 21, 57]. However, these papers largely have restricted themselves to simply noting the common features rather than explicitly studying whether adding these features leads to a more accurate model. Studying the transfer of data-mined models to new intelligent tutors is a highly important area of future work. So long as it is necessary to develop an entirely new model for each new intelligent tutoring system, the process of extending the advances in student modeling made through data mining to all tutoring systems will be slowed considerably. The methods from the data mining subfield of transfer learning [cf. 27, 28], transferring data-mined models to new contexts and new sampling distributions, may have a considerable amount to contribute to research in this area. More collaboration between transfer learning and educational data mining researchers would likely benefit both communities (providing a rich set of challenges to transfer learning researchers), as well as benefitting the developers of student models for intelligent tutoring systems.

A second limitation to educational data mining thus far is the lack of tools explicitly designed to support educational data mining research. Data mining tools such as Weka [59] and KEEL [1] support data mining practitioners in utilizing well-known data mining algorithms on their data, with a usable user interface; however, these tools' default user interfaces currently do not support the types of cross-validation that are necessary in educational data to infer generalizability across students. This has led to a considerable amount of EDM research that does not take these issues into account. It is possible to conduct cross-validation across data levels in another tool, RapidMiner [44]. RapidMiner does not directly support student-level or lesson-level cross-validation, but its "batch cross-validation" functionality makes it possible to conduct student-level or lesson-level cross-validation through pre-defining student batches outside of the data mining software. Beyond this issue, no data mining tool is currently integrated with tools for the text replay, survey, and quantitative field observation methods increasingly used to label data for using classification or regression, for student models. Integrating data mining tools with data labeling tools and providing support for conducting appropriate validation of generalizability would significantly facilitate research in using data mining to improve student models.

Even without these types of support, research into using data mining to support the development of student models has made significant impacts in recent years. To the degree that researchers address these limitations, the impact of educational data mining can be magnified still further.

7 Conclusion

In this chapter, we have discussed how data mining methods have contributed to the development of student models for intelligent tutoring systems. In particular, we have discussed the contribution to student modeling coming from classification methods, regression methods, clustering methods, and methods for the distillation of data for human judgment. Classification, regression, and clustering methods have supported the development of validated models of a variety of complex constructs that have been embedded into increasingly sophisticated student models, enabling broader-based adaptation to individual differences between students than was previously possible. Clustering methods have supported the discovery of how students choose to respond to new types of educational human-computer interactions, enriching student models; classification and regression models have afforded accurate and validated models of a broader range of student behavior. Among other constructs, these methods have supported the development of models of gaming the system, help-seeking, boredom, frustration, confusion, engaged concentration, self-efficacy, scientific reasoning strategies, and off-task behavior. Distillation of data for human judgment has itself facilitated the development of models of this nature, speeding the process of labeling data with reference to difference in student behaviors, in turn speeding the process of creating classification and regression models. In turn, these discoveries have increased the sophistication and richness of student models, covering a broader range of behavior. These richer student models have afforded more broad-based adaptation to student individual differences, significantly improving student learning.

Acknowledgements

The development of this paper was supported by the Pittsburgh Science of Learning Center (National Science Foundation) via grant "Toward a Decade of PSLC Research", award number SBE-0836012, by the National Science Foundation via grant "AMI: ASSISTments Meets Inquiry", award DRL#0733286, and by the U.S. Department of Education via grant "Promoting Robust Understanding of Genetics with a Cognitive Tutor that Integrates Conceptual Learning with Problem Solving", award number #R305A090549. All opinions presented in this documents represent the opinions of the author, and do not necessarily represent these opinions or viewpoints of any of the funders.

References

1. Alcala-Fdez, J., Sánchez, L., Garcıa, S., del Jesús, M., Ventura, S., Garrell, J., Otero, J., Romero, C., Bacardit, J., Rivas, V., Fernández, J., Herrera, F.: KEEL: A software tool to assess evolutionary algorithms for data mining problems, Soft Computing, vol. 13, no. 3, 307–318 (2008)

2. Aleven, V., McLaren, B.M., Roll, I., Koedinger, K.R.: Toward tutoring help seeking: Applying cognitive modeling to meta-cognitive skills. In: Proceedings of the 7th International Conference on Intelligent Tutoring Systems (ITS 2004), 227--239 (2004)
3. Aleven, V., McLaren, B., Roll, I., Koedinger, K.: Toward meta-cognitive tutoring: A model of help seeking with a Cognitive Tutor. International Journal of Artificial Intelligence and Education, 16, 101--128 (2006).
4. Amershi, S., Conati, C.: Combining Unsupervised and Supervised Classification to Build User Models for Exploratory Learning Environments. Journal of Educational Data Mining, 1 (1), 18--71 (2009).
5. Arroyo, I., Cooper, D., Burleson, W., Woolf, B. P., Muldner, K., Christopherson, R.: Emotion Sensors Go To School. In: Proceedings of the International Conference on Artificial Intelligence in Education, 17--24 (2009).
6. Baker, R.S.J.d.: Data Mining for Education. To appear in McGaw, B., Peterson, P., Baker, E. (Eds.) International Encyclopedia of Education (3rd edition). Oxford, UK: Elsevier. (in press)
7. Baker, R.S.J.d.: Modeling and Understanding Students' Off-Task Behavior in Intelligent Tutoring Systems. In: Proceedings of ACM CHI 2007: Computer-Human Interaction, 1059--1068 (2007).
8. Baker, R.S.J.d., Corbett, A.T., Aleven, V.: More Accurate Student Modeling Through Contextual Estimation of Slip and Guess Probabilities in Bayesian Knowledge Tracing. In: Proceedings of the 9th International Conference on Intelligent Tutoring Systems, 406--415 (2008)
9. Baker, R.S., Corbett, A.T., Koedinger, K.R.: Detecting Student Misuse of Intelligent Tutoring Systems. In: Proceedings of the 7th International Conference on Intelligent Tutoring Systems, 531--540 (2004).
10. Baker, R.S.J.d., Corbett, A.T., Koedinger, K.R., Evenson, S.E., Roll, I., Wagner, A.Z., Naim, M., Raspat, J., Baker, D.J., Beck, J.: Adapting to When Students Game an Intelligent Tutoring System. In: Proceedings of the 8th International Conference on Intelligent Tutoring Systems, 392--401 (2006).
11. Baker, R.S.J.d., Corbett, A.T., Roll, I., Koedinger, K.R.: Developing a Generalizable Detector of When Students Game the System. User Modeling and User-Adapted Interaction, 18, 3, 287--314 (2008).
12. Baker, R.S.J.d., Corbett, A.T., Wagner, A.Z.: Human Classification of Low-Fidelity Replays of Student Actions. In: Proceedings of the Educational Data Mining Workshop at the 8th International Conference on Intelligent Tutoring Systems, 29--36 (2006).
13. Baker, R.S.J.d., de Carvalho, A. M. J. A.: Labeling Student Behavior Faster and More Precisely with Text Replays. In: Proceedings of the 1st International Conference on Educational Data Mining, 38--47 (2008).
14. Baker, R.S.J.d., Mitrovic, A., Mathews, M.: Detecting Gaming the System in Constraint-Based Tutors. To appear in Proceedings of the 18th International Conference on User Modeling, Adaptation, and Personalization (in press).
15. Baker, R.S.J.d., Yacef, K.: The State of Educational Data Mining in 2009: A Review and Future Visions. Journal of Educational Data Mining, 1 (1), 3--17 (2009).
16. Beal, C.R., Qu, L., Lee, H.: Classifying learner engagement through integration of multiple data sources. In: Proceedings of the 21st National Conference on Artificial Intelligence, Boston, 2--8 (2006).
17. Beck, J.: Engagement tracing: using response times to model student disengagement. In: Proceedings of the 12th International Conference on Artificial Intelligence in Education (AIED 2005), 88--95 (2005).
18. Borgman, C. L., Abelson, H., Dirks, L., Johnson, R., Koedinger, K. R., Linn, M. C., Lynch, C. A., Oblinger, D. G., Pea, R. D., Salen, K., Smith, M. S., Szalay, A.: Fostering Learning in the Networked World: The Cyberlearning Opportunity and Challenge, A 21st Century

Agenda for the National Science Foundation. National Science Foundation: Washington, DC (2008).
19. Bradley, P.S., Fayyad, U., Reina, C.: Scaling EM (expectation maximization) clustering to large databases. Technical Report MSR-TR-98-35, Microsoft Research, Redmond,WA (1998).
20. Buckley, B. C., Gobert, J., Horwitz, P.: Using log files to track students' model-based inquiry. In: Proceedings of the 7th International Conference on Learning Sciences, 57--63 (2006).
21. Cetintas, S., Si, L., Xin,Y.P., Hord, C.: Automatic Detection of Off-Task Behaviors in Intelligent Tutoring Systems with Machine Learning Techniques, IEEE Transactions on Learning Technologies (2009)
22. Chaffar, S., Derbali, L., Frasson, C.: Inducing positive emotional state in Intelligent Tutoring Systems. In: Proceedings of the 14th International Conference on Artificial Intelligence in Education (2009)
23. Cocea, M., Hershkovitz, A., Baker, R.S.J.d.: The Impact of Off-task and Gaming Behaviors on Learning: Immediate or Aggregate? In: Proceedings of the 14th International Conference on Artificial Intelligence in Education, 507—514 (2009).
24. Cohen, J.: A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20 (1), 37--46 (1960).
25. Conati, C., Maclaren, H.**:** Empirically building and evaluating a probabilistic model of user affect. User Modeling and User-Adapted Interaction, 19 (3), 267**—**303 (2009).
26. Corbett, A.T., Anderson, J.R.: Knowledge Tracing: Modeling the Acquisition of Procedural Knowledge. User Modeling and User-Adapted Interaction 4, 253--278 (1995).
27. Dai, W.,Yang, Q., Xue, G., Yu, Y.: Boosting for transfer learning. In: Proceedings of the International Conference on Machine Learning (2007).
28. Daume, H.: Frustratingly easy domain adaptation. In the Proceedings of the Association for Computational Linguistics (2007).
29. de Vicente, A., Pain, H.: Informing the detection of the students' motivational state: an empirical study. In S. A. Cerri, G. Gouarderes, F. Paraguacu, editors, In: Proceedings of the Sixth International Conference on Intelligent Tutoring Systems, volume 2363 of Lecture Notes in Computer Science, 933--943, Springer (2002).
30. D'Mello, S. K., Craig, S.D., Witherspoon, A. W., McDaniel, B. T., Graesser, A. C.: Automatic Detection of Learner’s Affect from Conversational Cues. User Modeling and User-Adapted Interaction, 18(1-2), 45--80 (2008).
31. D’Mello, S. K., Picard, R., Graesser, A. C.: Towards an Affect Sensitive Auto-Tutor, IEEE Intelligent Systems, 22(4), 53--61 (2007).
32. Feigenbaum, E. A., McCorduck, P.: The fifth generation: Artificial Intelligence and Japan's Computer Challenge to the World. Addison-Wesley, Reading, MA (1983).
33. Grossberg, S.: Nonlinear neural networks: principles, mechanisms and architectures. Neural Networks 1:17--61 (1988).
34. Hartigan, J. A., Wong, M. A.: A K-means clustering algorithm. Applied Statistics, 28: 100–108 (1979).
35. Hawkins, D. M.: The Problem of Overfitting. Journal of Chemical Infomration and Modeling, 44, 1--12 (2004).
36. Hershkovitz, A., Nachmias, R.: Developing a Log-Based Motivation Measuring Tool. In: Proceedings of the First International Conference on Educational Data Mining, 226--233 (2008).
37. Johns, J., Woolf, B.: A Dynamic Mixture Model to Detect Student Motivation and Proficiency. In: Proceedings of the 21st National Conference on Artificial Intelligence (AAAI-06), 163--168 (2006).
38. Kay, J., Maisonneuve, N., Yacef, K., Reimann, P.: The Big Five and Visualisations of Team Work Activity. In: Proceedings of Intelligent Tutoring Systems (ITS06), M. Ikeda, K.

D. Ashley & T-W. Chan (eds). Lecture Notes in Computer Science 4053, Springer, 197--206 (2006).
39. Koedinger, K.R., Baker, R.S.J.d., Cunningham, K., Skogsholm, A., Leber, B., Stamper, J.: A Data Repository for the EDM community: The PSLC DataShop. To appear in Romero, C., Ventura, S., Pechenizkiy, M., Baker, R.S.J.d. (eds.) Handbook of Educational Data Mining. Boca Raton, FL: CRC Press. (in press)
40. Marder, M. Bansal, D.: Flow and diffusion of high-stakes test scores. In: Proceedings of the National Academy of Sciences of the United States, 106 (41), 17267--17270 (2009).
41. Maris, E.: Psychometric Latent Response Models. Psychometrika, 60 (4), 523--547 (1995).
42. McQuiggan, S., Mott, B., Lester, J.: Modeling Self-Efficacy in Intelligent Tutoring Systems: An Inductive Approach. User Modeling and User-Adapted Interaction 18, 81--123 (2008).
43. Merceron, A., Yacef, K.: Educational Data Mining: a Case Study. In: Proceedings of the International Conference on Artificial Intelligence in Education (AIED2005), IOS Press, 467--474 (2005).
44. Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., Euler, T.: YALE: Rapid Prototyping for Complex Data Mining Tasks. In: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2006). 935--940 (2006).
45. Montalvo, O., Baker, R.S.J.d., Sao Pedro, M., Nakama, A., Gobert, J.D.**:** Recognizing Student Planning with Machine Learning. Paper under review
46. Nkambou, R., Nguifo, E.M., Couturier, O., Fourier-Vigier, P.: Problem-Solving Knowledge Mining from Users' Actions in an Intelligent Tutoring System. LNCS 4509, 393--404 (2007).
47. Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann, San Francisco (1993).
48. Reynolds, D. A.: Gaussian Mixture Models, Encyclopedia of Biometric Recognition, Springer (2008).
49. Roll, I., Baker, R.S., Aleven, V., McLaren, B.M., Koedinger, K.R.: Modeling Students' Metacognitive Errors in Two Intelligent Tutoring Systems. In: Proceedings of User Modeling 2005, 379--388 (2005).
50. Romero, C., Ventura, S.: Educational Data Mining: A Survey from 1995 to 2005. Expert Systems with Applications 33, 125--146 (2007).
51. Sao Pedro, M., Baker, R.S.J.d., Montalvo, O., Nakama, A., Gobert, J.D.**:** Using Text Replay Tagging to Produce Detectors of Systematic Experimentation Behavior Patterns. Paper under review
52. Shih, B., Koedinger, K.R., Scheines, R.: A response time model for bottom-out hints as worked examples. In Baker, R.S.J.d., Barnes, T., & Beck, J.E. (Eds.) In: Proceedings of the First International Conference on Educational Data Mining, 117--126 (2008).
53. Steinwart, I., Christmann, A.: Support Vector Machines. Springer, Berlin (2008).
54. Studer, R., Benjamins, V.R., Fensel, D.: Knowledge engineering: Principles and methods. Data and Knowledge Engineering (DKE), 25(1–2):161–197 (1998).
55. Tang, T., McCalla, G.: Smart recommendation for an evolving e-learning system: architecture and experiment, International Journal on E-Learning, 4 (1), 105–129 (2005).
56. Teigen, K.H.: Yerkes-Dodson: a law for all seasons, Theory & Psychology, vol. 4, no. 4, 525–547 (1994).
57. Walonoski, J.A., Heffernan, N.T.: Detection and Analysis of Off-Task Gaming Behavior in Intelligent Tutoring Systems. In: Proceedings of the 8th International Conference on Intelligent Tutoring Systems, 382--391 (2006).
58. Walonoski, J.A., Heffernan, N.T. Prevention of Off-Task Gaming Behavior in Intelligent Tutoring Systems. In: Proceedings of the 8th International Conference on Intelligent Tutoring Systems, 722--724 (2006).
59. Witten, I.H., Frank, E.: Data mining: Practical Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufmann, San Francisco, CA (1999).

60. Woolf, B.P., Arroyo, I., Muldner, K., Burleson, W., Cooper, D., Dolan, R., Christopherson, R.M.: The Effect of Motivational Learning Companions on Low-Achieving Students and Students with Learning Disabilities. To appear in the Proceedings of the International Conference on Intelligent Tutoring Systems (in press).
61. Yu, L., Liu, H.: Feature selection for high-dimensional data: a fast correlation-based filter solution. In: Proceedings of the International Conference on Machine Learning, 856--863 (2003).
62. Zaïane, O.: Building a recommender agent for e-learning systems. In: Proceedings of the International Conference on Computers in Education, 55–59 (2002).