Learning, Moment-by-Moment and Over the Long Term

Yang Jiang[1], Ryan S. Baker[1], Luc Paquette[1], Maria San Pedro[1], Neil T. Heffernan[2]

[1] Teachers College, Columbia University, New York, NY, United States
[2] Worcester Polytechnic Institute, Worcester, MA, United States

yj2211@tc.columbia.edu, baker2@exchange.tc.columbia.edu, paquette@tc.columbia.edu, mzs2106@tc.columbia.edu, nth@wpi.edu

Abstract. The development of moment-by-moment learning graphs (MBMLGs), which plot predictions about the probability that a student learned a skill at a specific time, has already helped to improve our understanding of how student performance during the learning process relates to robust learning [1]. In this study, we extend this work to study year-end learning outcomes and to account for differences in learning on original questions and within knowledge-construction scaffolds. We discuss which quantitative features of moment-by-moment learning in these two contexts are predictive of the longer-term outcomes, and conclude with potential implications for instruction.

Keywords: Moment-by-moment learning, scaffolding, intelligent tutoring system, educational data mining.

1 Introduction

Recent advancements in educational data mining have allowed researchers to infer the probability that a student learned a skill at a specific time during learning [2]. With these estimates, it becomes possible to construct visual graphs of individual students' learning over time (moment-by-moment learning graphs, or MBMLGs). Different visual patterns of MBMLGs obtained from student usage of a tutor are associated with differences in student learning outcomes [3]. However, this type of visual analysis requires a human analyst. In more recent work, Hershkovitz and colleagues [1] distilled quantitative features from MBMLGs, in order to predict robust learning. In current study, we extend this work to analyze whether the features can predict a longer-term outcome, standardized exam performance. Another question addressed in this paper is whether learning in different contexts impacts these patterns, in particular whether there is a difference between original questions and scaffolding questions.

To research these questions, we replicate the quantitative features of MBMLGs that were successful at predicting robust learning outcomes with reasonable precision [1] and extend this prior work by distinguishing between original questions and scaffolding questions. We then analyze how these features of students' MBML correlate to student performance on the Massachusetts Comprehensive Assessment System (MCAS), a high-stakes standardized test given at the end of the year. We compute the

adfa, p. 1, 2011.

correlations to outcomes for each of the features, specifically comparing the differences in correlations for original questions versus scaffolding questions.

2 Moment by Moment Learning Graph Features

We study the questions within the context of ASSISTments [4], a web-based tutoring system for middle school mathematics. ASSISTments data were used to investigate the correspondence between the fine-grained quantitative attributes of the MBMLGs and student math performance on the MCAS. This was done in two stages.

First, MBMLGs were constructed using a machine-learned model of MBML, using data from 7,647 middle school students from four school districts who used ASSISTments throughout an entire school year (2004-2005 to 2008-2009). Overall, students completed a total of 2,281,808 actions (i.e., submitting an answer or requesting help) across a range of 19,991 problems within the system. Next, a discovery with models approach was used to explore the relationship between MBML and MCAS scores among a subset of 613 students in one urban district for whom MCAS scores were available. Students used ASSISTments in the classroom as preparation for the MCAS test for two hours, twice a week, throughout the 2004-2005 school year, completing a total of 97,245 actions (56,343 on original questions and 40,902 on scaffolding questions) targeting a broad range of mathematical skills.

The construction of MBMLGs is a three-step process, described in detail in [2]. First, each problem step in the data is labeled with the probability *P(J)* that the student learned that skill on that particular attempt, using data from the student's future performance. Second, a machine-learned model that predicts *P(J)* is built from a broad set of features, using data only from the student's past and present performance. Last, we integrate across predictions to construct a MBMLG for each student/skill.

Once MBMLGs were created, a feature set (see Table 1) was distilled from the quantitative characteristics of each graph. In order to account for differences between original and scaffolding questions, we computed the MBMLG features separately for the original questions (denoted *o*) and scaffolding questions (denoted *s*). For each student, MBMLG features (except *sumByLen* and *areaByLen*) were computed separately for each skill, and averaged across all skills.

Table 1. List of all the features distilled from the MBMLGs.

***avgMBML**: Average moment-by-moment learning value in a given graph.*
***sumMBML**: Sum of moment-by-moment learning values in a given graph.*
***graphLen**: Number of steps in a MBMLG (number of problems received).*
***area**: Area under the MBMLG.*
***peak**: Height of the largest peak in the MBMLG.*
***2ndPeak**: Height of the 2nd-largest peak in the MBMLG.*
***3rdPeak**: Height of the 3rd-largest peak in the MBMLG.*
***peakIndex**: First index of the largest peak in the MBMLG (Index = 1 equals the first step involving the skill).*
***2ndPeakIndex**: First index of the 2nd-largest peak in the MBMLG.*
***2PeakDist**: Distance between the largest and the 2nd-largest peaks.*
***2PeakAdjDist**: 2PeakDist, divided by graphLen.*
***2PeakDecr**: Decrease [%] of magnitude from largest to 2nd-largest peak.*
***2PeakDist-adjDecr**: 2PeakDecr divided by 2PeakDist.*
***3PeakDecr**: Decrease [%] of magnitude from largest to 3rd-largest peak.*
***3PeakDist-adjDecr**: 3PeakDecr divided by 3PeakDist.*
***sumByLen**: Avg. sumMBML across skills for student divided by avg. graphLen for that student.*
***areaByLen**: Avg. area for student across skills divided by average graphLen for that student.*

3 The Relationships between Individual Features of the MBMLGs and Long-Term Learning Outcomes

In this section, we explore the relationships between individual features of the MBMLGs (defined in Section 2) and student math scores on MCAS, using correlation mining and significance tests with post-hoc controls (Storey's q-value method [5]), and whether the features based on original or scaffolding questions better predicted the MCAS, using statistical tests of the difference between two correlation coefficients for correlated samples with post-hoc controls.

Table 2. Correlation of MBMLG features to MCAS scores. ρ_{OM} and ρ_{SM} denote the Spearman correlation between MCAS scores and MBMLG features for *original* and *scaffolding* questions, respectively. Correlations that are sig. after controlling for false discovery (q<0.05) are marked by *.

Feature	ρ_{OM}	q_{OM}	ρ_{SM}	q_{SM}	t	q
avgMBML	-0.180	<0.001*	-0.046	0.080	-3.559	<0.001*
sumMBML	0.086	0.012*	-0.219	<0.001*	7.793	<0.001*
graphLen	0.275	<0.001*	-0.103	0.004*	9.632	<0.001*
area	0.130	<0.001*	-0.218	<0.001*	8.829	<0.001*
peak	0.031	0.136	0.082	0.015*	-1.273	0.070
2ndPeak	-0.216	<0.001*	-0.188	<0.001*	-0.652	0.152
3rdPeak	-0.283	<0.001*	-0.263	<0.001*	-0.435	0.188
peakIndex	0.112	0.002*	-0.216	<0.001*	7.775	<0.001*
2ndPeakIndex	0.140	<0.001*	-0.090	0.011*	4.543	<0.001*
2PeaksDist	0.145	<0.001*	-0.124	0.001*	5.379	<0.001*
2PeakAdjDist	-0.227	<0.001*	-0.037	0.113	-3.548	<0.001*
2PeakDecr	0.325	<0.001*	0.480	<0.001*	-3.875	<0.001*
2PeakDist-adjDecr	0.307	<0.001*	0.465	<0.001*	-3.804	<0.001*
3PeakDecr	0.323	<0.001*	0.477	<0.001*	-3.630	<0.001*
3PeakDist-adjDecr	0.150	<0.001*	0.381	<0.001*	-4.443	<0.001*
sumByLen	-0.300	<0.001*	-0.147	<0.001*	-4.150	<0.001*
areaByLen	-0.081	0.015*	-0.276	<0.001*	4.476	<0.001*

Table 2 shows the Spearman's correlations ρ between individual features and MCAS scores, and their post-hoc controlled statistical significance q. The strongest correlations involved differences between the largest peak values during scaffolding, including *2PeakDecr (s)*, the decrease [%] in magnitude from the largest to 2nd-largest peak on scaffolding questions ($\rho_{SM}(585) = 0.480$, $q < 0.001$); and *3PeakDecr (s)*, the decrease [%] in magnitude from the largest to 3rd-largest peak on scaffolding questions ($\rho_{SM}(546) = 0.477$, $q < 0.001$). Larger differences between these values indicate "spikier" graphs where considerable learning occurs in eureka learning moment(s) [1]. A weaker version of the same pattern was found for original questions.

MCAS scores were positively correlated with *peakIndex (o)*, the index of the largest peak in the Original MBMLG ($\rho_{OM}(612) = 0.112$, $q = 0.002$), but negatively correlated with *peakIndex (s)*, the index of the largest peak in the Scaffolding MBMLG ($\rho_{SM}(612) = -0.216$, $q < 0.001$). Most likely, this difference (which was significant, $t(609) = 7.775$, $q < 0.001$) demonstrates the contribution that the ASSISTments scaffolding system makes to learning. If students had their highest learning late in the learning process involving original questions (possibly due to scaffolding beforehand), they did better on the exam. But for scaffolding questions, earlier moments of high learning were associated with higher MCAS scores.

Other results also confirmed the importance of the learning context. Answering more original questions in ASSISTments (*graphLen)* was associated with higher test performance ($\rho_{OM}(612) = 0.275$, $q < 0.001$). In contrast, more problem steps on scaffolding questions corresponded to poorer learning outcomes ($\rho_{SM}(612) = -0.103$, $q = 0.004$). This difference ($t(609) = 9.632$, $q < 0.001$) likely reflects the fact that students receive more scaffolding if they are performing poorly. Similarly, area under the Original MBMLG (*area (o)*) was positively correlated with the MCAS while area under the Scaffolding MBMLG (*area (s)*) was negatively associated with the MCAS, the difference between the correlations ($t(609) = 8.829$, $q < 0.001$) was significant.

4 Conclusion

This paper explores the relationships between quantitative features of MBMLGs and students' performance on an end-of-year exam, comparing features based on performance during original questions and scaffolded tutoring. This separation allows us to discover significant temporal effects on student learning; students who demonstrate high learning early in their interactions with ASSISTments through scaffolding are most likely to perform well on the state exam. This finding suggests that the scaffolding in ASSISTments may be useful beyond simply producing better performance on the current skill. In general, the MBMLG appears to be able to shed light on fine-grained aspects of the learning process that are associated with important outcomes; figuring out the best uses of this method is an area for further research.

5 Acknowledgments

This research was supported by the National Science Foundation # SBE-0836012.

References

1. Hershkovitz, A., Baker, R.S.J.d., Gowda, S.M., Corbett, A.T.: Predicting Future Learning Better Using Quantitative Analysis of Moment-by-Moment Learning. Proceedings of the 6th International Conference on Educational Data Mining, 74-81 (2013)
2. Baker, R.S.J.d., Goldstein, A.B., Heffernan, N.T.: Detecting Learning Moment-by-Moment. IJAIED, 21 (1-2), 5-25 (2011)
3. Baker, R.S.J.d., Hershkovitz, A., Rossi, L.M., Goldstein, A.B., & Gowda, S.M.: Predicting Robust Learning with the Visual Form of the Moment-by-Moment Learning Curve. Journal of the Learning Sciences, 22 (4), 639-666 (2013)
4. Razzaq, L., et al.: Blending Assessment and Instructional Assistance. In: Nedjah, N., de Macedo Mourelle, L., Borges, M.N., and Almeida, N.N. (eds.) Intelligent Educational Machines, 23-49. Springer, Berlin / Heidelberg (2007)
5. Storey J.D., Taylor J.E., & Siegmund D.: Strong Control, Conservative Point Estimation, and Simultaneous Conservative Consistency of False Discovery Rates: A Unified Approach. Journal of the Royal Statistical Society, Series B, 66, 187-205 (2004)