

# Learner-Context Modelling: A Bayesian Approach

Charles  $Lang^{(\boxtimes)}$ 

Teachers College, Columbia University, New York, NY 10027, USA charles.lang@tc.columbia.edu

Abstract. The following paper is a proof-of-concept demonstration of a novel Bayesian model for making inferences about individual learners and the context in which they are learning. This model has implications for both efforts to create rich open leaner models, develop automated personalization and increase the breadth of adaptive responses that machines are capable of. The purpose of the following work is to demonstrate, using both simulated data and a benchmark dataset, that the model can perform comparably to commonly used models. Since the model has fewer parameters and a flexible interpretation, comparable performance opens the possibility of utilizing it to extend automation greater variety of learning environments and use cases.

Keywords: Context modelling  $\cdot$  Personalization  $\cdot$  Individualization  $\cdot$  Open learner model  $\cdot$  Bayes

# 1 Introduction

## 1.1 Learner-Context Models

The growth of artificial intelligence in education will be determined to some extent by our ability to expand into new formats and data collection contexts and of machines to model the learner across these disparate environments [9,11]. Here we take a tentative step towards an extensible learner model that would allow individual learner modelling across many different contexts and task types, as well as content domains. We build on work to create a Bayesian Learner-Context model that can support a wide range of task formats [6,7] and provide an alternative to other context modelling attempts [1,3,8]. The purpose of this paper is to introduce the model and benchmark it against other models with respect to prediction accuracy. Based on the results presented here we believe the model can find utility in expanding automated responses due to its simpler parameterization and more flexible interpretation. The research questions we intend to answer are:

- 1. How does the model perform on simulated data with respect to the recover of exact values? (i.e. if we knew the exact thoughts of learners)
- 2. How does the model compare to other models based on performance on benchmark data sets?

© Springer Nature Switzerland AG 2020

I. I. Bittencourt et al. (Eds.): AIED 2020, LNAI 12164, pp. 152–156, 2020. https://doi.org/10.1007/978-3-030-52240-7\_28

# 2 Methods

To capture the relationship between internal and external random variables, we appeal to Bayes rule, construing the internal factors as the learner's *prior* knowledge and the external factors as the likelihood of the context given the learner's belief. A Bayesian learner would take input from her environment to calculate a posterior probability of the truth of a hypothesis, given the current that environment (P(H|D)), from the likelihood of the data in light of the hypothesis (P(D|H)) and their prior belief in the hypothesis from their accumulated experience (P(H)) [4]. The likelihood is the degree to which the data confirms or dis-confirms the learner's belief in the hypothesis. The modeller's job then becomes to generate estimates of each individuals' likelihood and prior, to best predict their individual behavior at a task represented by the posterior probability. Within this framework, if we can characterize probabilistically a learner's prior knowledge and how that learner interprets their conditions we should be able to accurately predict their behavior. In other words, modelling learner behavior becomes a matter of resolving what each individual learner brings to the table vs. what the table brings to each learner:

#### $Behavior \propto Context \times PriorKnowledge \tag{1}$

The likelihood is what gives this model its ability to cover many different contexts, as long as the contexts can be coded, a probability distribution can be fit to them for each individual learner, represented as the Inverse Bayes Rule [10]:

$$\theta = \frac{\beta - \alpha\beta}{\alpha - 2\alpha\beta + \beta} \tag{2}$$

where  $\theta$  is the posterior probability,  $\alpha$  is the number of times the learner was correct and had experienced the specific context and  $\theta$  is the prior probability. Code, data and further explanation is available in the following GitHub Repository.

#### 2.1 Data

The data set used for analysis consists of 8,09, 12–14 year olds in the eighth grade of a school district in the North East of the United States during the 2009–10 school year. Student data were collected through ASSISTments, a web-based math tutoring system designed to prepare students for state standardized tests. Data consist of 603,128 log records. Each record is comprised of a timestamp recording when the learner answered the item, an item ID, student ID, the student's answer, the skill (of 153 possible skills) the item was testing and the type of item: multiple choice question, algebraic equation or text answer. All data was retrieved from ASSISTments [5]. No students can be identified.

## 3 Results

Figure 1A demonstrates certainty as we increase the number of conditions that the model is attempting to resolve while holding the number of hypotheses constant. Figure 1B demonstrates the reduction in error (represented by RMSE) of the estimate of the prior value over the number of items. For a single skill or hypothesis the estimate reaches within 0.1 of the true value within ten items.



Fig. 1. Simulation results showing the relationship between number of items and the accuracy of belief estimation (A), decreasing error rates estimating prior probabilities over sequences of answers (B).

A comparison to other prediction algorithms on a benchmark prediction task can provide an idea of the relative efficacy of this model. Pardos and Heffernan 2011 have published performance of Bayesian Knowledge Tracing (both individualized and standard) on this same ASSISTments data set with an average cross validated AUC of 0.67 and 0.69 respectively [8]. The context-learner model achieves a cross-validated average AUC of 0.64.

# 4 Discussion

This paper presents a novel algorithm for predicting learner actions within automated systems, building on previous work that characterized learners as Bayesians [6,7]. The method involves making predictions about individual learners using the sequence of actions and the environment that they are operating in. We further quantified how successful the model is at forecasting learner scores using simulated learner data and a benchmark data set drawn from the ASSISTments online tutoring system.

Since the data is manufactured we have the control to measure how well the model can infer the true belief of the learner. This is not possible in reality but is informative to understand some characteristics of the model and infer what may happen when it is applied. Of particular concern is whether or not the model has useful statistical power, in other words, whether its ability to estimate the learner's state across a given number of skills and conditions given the number of items they have attempted. Whether these estimates have the requisite statistical power to be of use is an interesting and open question that will take more study. What is certain is that it is context driven, whether 100 items is too onerous for the learner depends a lot on what the item is. If it is moves within a game this may not be onerous at all, if each item is a essay question, collecting 100 over a short period of time may well be unrealistic. In the simulation run here it appeared that on average a single skill could be estimated within 0.1 of the true prior value within 15 items.

The validation model demonstrates some interesting characteristics of the method. In opposition to the findings in the simulated data, here there does not seem to be a strong relationship between accuracy and the number of times a skill is tested. Skills with greater number of items devoted to them do not see greater prediction accuracy than those with fewer. One possible reason for this are that all skills had sufficient attempts so there was no observable effect. But there were observable differences across contexts. Different item contexts appear to have different false negative rates. The model does better at predicting the answers to multiple choice questions than text based answers, with text based answers having a higher false negative rate. That we can differentiate contexts according to their accuracy rates suggests that contexts can be parsed by this model to categorize learners. This may provide characterizations that could be used to inform computer decision making. A model like BKT is limited to the insights it can gain to its four parameters - knowing, demonstrating, slipping, and guessing [2]. This model, although having fewer parameters, can provide information across an infinite number of contextual factors because the parameters refer directly to both the learner and the learner's context.

There are three chief benefits of this model. 1. It expands the vocabulary of outcomes that can be quantified beyond things that can be classified as correct/incorrect to anywhere any situation in any behavioral change can be quantified. 2. It allows a distinction to be made between learner proficiency and the impact of the environment that the learner finds herself within and 3. It is an individualised measure that is defined absent reference to other learners so can support flexible, bottom-up analysis of groups. There is currently no method with these characteristics available and it may prove a useful addition to the analytic methodology as it allows us to make more efficacious statements about individual learners, rather than relying on subgroup allocation. The benefits for automated personalization are substantial, but also for context modelling as this is an essential part of the methodology. Since the model requires context to be numerically estimated, context cannot be ignored nor treated as noise.

## References

 Baker, R.S.J., Corbett, A.T., Aleven, V.: More accurate student modeling through contextual estimation of slip and guess probabilities in bayesian knowledge tracing. In: Woolf, B.P., Aïmeur, E., Nkambou, R., Lajoie, S. (eds.) Intelligent Tutoring Systems. pp. 406–415. Lecture Notes in Computer Science, Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-69132-7\_44

- Corbett, A.T., Anderson, J.R.: Knowledge tracing: modeling the acquisition of procedural knowledge. User Model. User-Adapted Interact. 4(4), 253–278 (1994)
- Fancsali, S.E., Ritter, S.: Context personalization, preferences, and performance in an intelligent tutoring system for middle school mathematics. In: Proceedings of the Fourth International Conference on Learning Analytics and Knowledge, pp. 73–77 (2014)
- Gopnik, A., Tenenbaum, J.: Bayesian networks, Bayesian learning and cognitive development. Dev. Sci. 10(3), 281–287 (2007). https://doi.org/10.1111/j.1467-7687.2007.00584.x
- Heffernan, N.T., Heffernan, C.L.: The ASSISTments ecosystem: building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. Int. J. Artif. Intell. Educ. 24(4), 470–497 (2014). https://doi.org/10.1007/s40593-014-0024-x
- Lang, C.: An adaptive model of student performance using inverse bayes. J. Learn. Anal. 1(3), 154–156 (2014)
- Lang, C.: Opportunities for personalization in modeling students as bayesian learners. In: Proceedings of the Seventh International Learning Analytics & Knowledge Conference, pp. 41–45 (2017)
- Pardos, Z.A., Heffernan, N.T.: KT-IDEM: introducing item difficulty to the knowledge tracing model. In: Konstan, J.A., Conejo, R., Marzo, J.L., Oliver, N. (eds.) User Modeling, Adaption and Personalization, pp. 243–254. Lecture Notes in Computer Science, Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-22362-4\_21
- Roll, I., Wylie, R.: Evolution and revolution in artificial intelligence in education. Int. J. Artif. Intell. Educ. 26(2), 582–599 (2016). https://doi.org/10.1007/s40593-016-0110-3
- Tian, G.L., Tan, M.: Exact statistical solutions using the Inverse Bayes Formulae. Stat. Probab. Lett. 62(3), 305–315 (2003). https://doi.org/10.1016/S0167-7152(03)00044-0
- Timms, M.J.: Letting artificial intelligence in education out of the box: educational cobots and smart classrooms. Int. J. Artif. Intell. Educ. 26(2), 701–712 (2016). https://doi.org/10.1007/s40593-016-0095-y