# Individualization of Bayesian Knowledge Tracing Through Elo-infusion.

Michael Yudelson[1] [0000−0002−8280−0348]

Yudelson Consulting LLC, Pittsburgh, PA 15217, USA
myudelson@gmail..com
http://www.yudelson.info

**Abstract** For as long as the Bayesian Knowledge Tracing (BKT) approach is known, so are the attempts to account for not only skill-level but individual student factors. A lot of computational methods to implement individualization in BKT were proposed over the past 25 years as BKT existed. To this day, virtually all individualization approaches were not suited for easy implementation. Either they were purely analytical (only fit for post-hoc analyses) or required significant computational effort to realize (e.g., calibrating individual factors as students cleared units of content).
In this work, we discuss implementing the individualization of BKT using a mechanism of an Elo rating schema. Elo has been established in the educational domain for some time and offers tangible theoretical and practical benefits. We show that infusing BKT even with an Elo component using a single parameter to track student-specific factors results in significant quantitative and qualitative improvements to modeling student learning. This approach is easy to implement in a system already featuring BKT.

**Keywords:** Bayesian Knowledge Tracing · Elo rating schema

## 1 Introduction

Bayesian Knowledge Tracing (BKT) [4] is one of the most researched approaches to tracking student learning in computer-assisted problem-solving applications. One of the popular thrusts of BKT-centric research is accounting for student-specific factors. Standard BKT contains only skill-specific parameters, although, admittedly, these parameters are population parameters. The issue of individualization first raised in the original BKT paper targets the variability in how students learn and perform above and beyond what is captured by skill-specific components of the model. Despite a significant volume of publications on the topic of individualizing BKT, the resulting approaches are largely analytical due to computational, implementational, and other considerations. This is why individualization of BKT was largely an analytical topic and, to the best of our knowledge, was never deployed in a setting other than experimental.

In this work, we are proposing an approach that, while remaining in the stream of iBKT research, is first and foremost operationalizable. Our suggestion is easy to implement in any new product suitable to be equipped with a standard BKT and easy to be added to a product already using standard BKT as its core student modeling method. We call our approach individualization through Elo-infusion or simply Elo-infusion. Elo is a family of rating schema methods used in multiple contexts from rating players in competitive sports to online dating and education. We are borrowing Elo's mechanism of tracking, in this case, student proficiency while solving problems and infusing a standard BKT with it.

## 2 Bayesian Knowledge Tracing and Its Individualization

There are four types of model parameters used in Bayesian Knowledge Tracing: an initial probability of knowing the skill a priori – $p(L_0)$ (or *p-init*), probability of student's knowledge of a skill transitioning from *not known* to a *known* state after an opportunity to apply it – $pT$ (or *p-transit*), the probability to make a mistake when applying a known skill – $pS$ (or *p-slip*), and the probability of correctly applying a not-known skill – $pG$ (or *p-guess*).

One notable individualization approach proposes to split BKT parameters to per-skill and to-per student components. Student and skill components are added in log-odds space and transformed back to probability space: see Equation 1a. Here, $w$ is one of the BKT parameters being individualized, $w_k$ is the per-skill component of the BKT parameter, and $w_i$ is the per-student component. This iBKT model is fit using a coordinate gradient descent method [6].

$$w = \sigma(logit(w_k) + logit(w_i)) \tag{1a}$$

$$\sigma(x) = 1/(1 + e^{-x}) \tag{1b}$$

$$logit(y) = ln(y/(1 - y)) \tag{1c}$$

## 3 Elo Rating Schema

Elo rating schema was proposed by Arpad Elo [1] and was originally used to rate chess players. Recently, it's been successfully used for tracking learners performance. Pelánek et al. use Elo to track knowledge of Geography [2]. Math Garden [3] deployed in K-12 setting in the Netherlands is based on Elo too.

Elo capturing students and items is shown in Equation 2a. Here, $p_{ij}$ – is the probability student answers the item correctly, $m_{ij}$ – is the log-odds value of that probability, $s_i$ – is student's unidimensional ability (initially 0), and $b_j$ – is question/problem unidimensional difficulty (initially 0). Tracked Elo values are updated as new data points are observed according to Equations 2b-2c. $K$ is a sensitivity parameter controlling the magnitude of the update.

$$p_{ij} = Pr(X_{ij} = 1) = \sigma(m_{ij}) = 1/(1 + e^{-m_{ij}}) = 1\Big/(1 + e^{-(s_i - b_j)}) \tag{2a}$$

$$s_i = s_i + K \cdot (X_{ij} - p_{ij}) \tag{2b}$$

$$b_j = b_j - K \cdot (X_{ij} - p_{ij}) \tag{2c}$$

## 4 Elo-infusion

We have devised an individualized Elo-infused BKT by combining per-student and per-skill components like in an iBKT approach featured in [6] and shown in Equation 1a. The per-skill parts of parameters ($w_k$) remain the same, while the per-student part ($s_i$) is taken from Elo (rf. Equation 3). The $sgn(w)$ function maps every individualized BKT parameter to $\{-1, 1\}$. For the $pL_0$ (prior mastery), $(1-pF) = 1$ (not forgetting), $pT$ (learning), $1-pS$ (not slipping), $pG$ (guessing), $sgn(w) = 1$ . One could commonly call this group *positive effects on performance*. For the remaining vector-matrix parameters $sgn(w) = -1$. Running values of the parameters are updated according to standard BKT or Elo rules. As a result, we have four BKT parameters per skill and one parameter for updating all student ratings. Just like in the iBKT approach by Yudelson and colleagues [6], different subsets of BKT parameters could be the target of infusion. One could infuse all – priors (p-init), transitions (p-learn), and emissions (p-slip and p-guess), or any combination of the three groups.

$$w = \sigma(sgn(w) \cdot s_i + logit(w_k)) \tag{3}$$

## 5 Data

We used the datasets from the KDD Cup 2010[1]. The data was contributed by Carnegie Learning Inc., a publisher of math curricula and a producer of intelligent tutoring systems for middle school and high school. There are two datasets, Algebra I, and Bridge to Algebra, both collected in the 2008-2009 school year. We removed the rows that had no skill tagging. Just like in the original Cognitive Tutor, skills are treated as unique within sections of math content. We obtained original BKT parameters shipped with the Cognitive Tutor product from Carnegie Learning, Inc.

## 6 Computational Experiments

We implemented an Elo-infused BKT model described above based on the `hmm-scalable` [6]. To test the new approach, we used the datasets described above to compare it to the standard BKT using parameters shipped with the original Cognitive Tutor as well as to standard BKT and iBKT models fit to the data.

The task of the gradient-based search for sensitivity parameter $K$ was simplified by enumerating candidate sensitivities from 0.0001 to 1.0 using a factor of 5 and 2 (yielding values ending in 5 and 1). The fitting of the BKT part of the model remains computationally correct even after introducing the Elo-infusion. Different parameter scopes were targeted. Namely, just p-init (Pi), p-init and

[1] KDD Cup 2010 Educational Datamining Challenge http://pslcdatashop.web.cmu.edu/KDDCup

Table 1: Best infused models, their respective scopes, and Elo sensitivities $K$ compared to reference models.

| Dataset | Model | Infusion scope | Sensitivity | Accuracy | RMSE |
|---|---|---|---|---|---|
| Algebra I | Shipped BKT | | | 0.7557 | 0.4367 |
| Algebra I | Fit BKT | | | 0.8304 | 0.3566 |
| Algebra I | Elo-infused fit BKT | Pi,A,B | 0.030 | 0.8325 | 0.3532 |
| Algebra I | Elo-infused shipped BKT | Pi,A,B | 0.300 | 0.8200 | 0.3731 |
| B. to Algebra | Shipped BKT | | | 0.7994 | 0.3840 |
| B. to Algebra | Fit BKT | | | 0.8333 | 0.3516 |
| B. to Algebra | Elo-infused fit BKT | Pi,A | 0.010 | 0.8351 | 0.3494 |
| B. to Algebra | Elo-infused shipped BKT | Pi,A,B | 0.300 | 0.8234 | 0.3624 |

p-learn (Pi,A), just p-learn (A), and all parameters p-init, p-learn, p-slip, and p-guess (Pi,A,B).

To draw comparisons between the alternative models, we used a combined 5 times 2-fold student-stratified cross-validation F-test to compare models [5]. This approach was validated on multiple datasets and shown reliable model ranking results. The use of 2- fold cross-validation defends against increased overlap of the training sets when the number of folds is 3 or more. This approach has a low Type I error rate.

In terms of model performance metrics, we used accuracy and root mean squared error (RMSE). Accuracy lets us know how often the model predicts the right answer outcome (right or wrong). RMSE tells us how far numerically from the correct outcome our prediction was. For each metric, we computed an aggregated mean value across 10 training-prediction rounds. Whenever necessary, we applied the 5x2-fold F-test to obtain a significance value (p-value) for every pair of the models compared.

## 7 Results

The computational cross-validation experiments are summarized in table Table 1. There, we give averages over 10 prediction tasks (5 random runs of 2-fold validation). For each of the considered datasets the lowest performing reference point is the shipped model. The best performing model is the Elo-infused fit BKT. It is worth noting, that the Elo-infused shipped BKT (fixed BKT skill parameters but) is a significant improvement over the shipped model. Elo-infused shipped BKT is about half-way between the reference shipped BKT and the best Elo-infused fit BKT in both datasets. In terms of pairwise comparisons of accuracy/RMSE model performances – all are statistically significantly different even if correction to account for multiple comparisons are made.

## References

1. Elo, A.E.: The rating of chessplayers, past and present. Arco Publishers (1978)

2. Nižnan, J. and Pelánek, R. and Rihák, J.: Student models for prior knowledge estimation. In: Proceedings of the 8th International Conference on Educational Data Mining. pp. 109–116. ACM, New York, NY, USA (2015)
3. Hofman, A., Jansen, B., de Mooij, S., Stevenson, C., van der Maas, H.: A solution to the measurement problem in the idiographic approach using computer adaptive practicing. Journal of Intelligence **6**(1), 14 (2018)
4. Corbett, A. T. and Anderson, J. R.: Knowledge tracing: Modeling the acquisition of procedural knowledge. User Modeling and User-Adapted Interaction, **4**(4), 253–278 (1995)
5. Alpaydm, E.: Combined 5x2 cv F-test for comparing supervised classification learning algorithms. Neural computation **11**(8), 1885–1892. (1999) Author, F.: Article title. Journal **2**(5), 99–110 (2016)
6. Yudelson, M., Koedinger, K., Gordon, G.: Individualized Bayesian Knowledge Tracing Models. In: 16th International Conference on Artificial Intelligence in Education (AIED 2013), pp. 171–180. Springer, Memphis, TN (2013).