Answer Correctness Prediction

Tomáš Mazel

ČVUT - FIT

mazeltom@fit.cvut.cz

December 31, 2020

1 Introduction

Knowledge Tracing (KT), i.e. tracing student knowledge from their learning history, is one of the most important problems in the field of Artificial Intelligence in Education (AIEd) [18, 9, 11, 15, 24]. Through KT, an Intelligent Tutoring System (ITS)[3] can understand each student's learning behavior and provide learning experience adapted to all individuals [13].

A variety of methods including Bayesian Knowledge Tracing (BKT)[23, 17, 22, 6], Deep Knowledge Tracing (DKT) [18], various reinforcement learning techniques [19, 12, 8, 14, 20, 25, 21] and many more [15, 2, 24, 16] have been developed. The question which methods are most effective for KT is still unsettled, partially because of limited scale of available datasets [7, 16, 10, 4, 9, 11].

In order to help answer this question the authors of this task [1] have developed a large-scale ITS dataset called EdNet with different student activities ranging from question solving to lecture watching activities [5]. The EdNet dataset is the largest of its kind in the world, containing 123M interactions coming from more than 1M users.

The aim of the project [1] is to perform the following tasks with the EdNet data:

1. Correctness Prediction

Input: Each user's history of learning behaviors (question responses and lecture watching activities) given in chronological order
Output: Expected correctness probability (correct/incorrect) for each newly encountered question
Goal: To predict a student's response correctness (correct/incorrect) to newly encountered multiple-choice questions assessing certain parts of their English skills.
Metrics: Accuracy, ROC-AUC

2. Response Prediction

Input: Each user's history of learning behaviors (question responses and lecture watching activities) given in chronological order
Output: Expected response (option a,b,c or d) for each newly encountered question
Goal: To predict a student's response to newly encountered multiple-choice questions assessing certain parts of their English skills.
Metrics: Accuracy, ROC-AUC

3. Dropout Prediction

Input: Each user's history of learning behaviors (question responses and lecture watching activities) given in chronological order
Output: Expected dropout probability for each newly encountered activity (question, lecture)
Goal: To predict a student's likelihood for dropout (large gap in learning activities) during learning activities.
Metrics: Accuracy, ROC-AUC

2 Dataset Description

The EdNet dataset contains complete lists of interactions students made on SANTA [8], an ITS specifically developed for preparing students for the TOEIC exam. For each student, the interactions are sorted and provided in a separate table. Since the table includes educational tags of each learning item, methods like BKT can effectively make use of pedagogical properties to estimate a student's knowledge state.

Examples of training data, example test, lectures and questions are available at: `https://gitlab.fit.cvut.cz/mazeltom/mvi-sp/blob/master/Examples_of_available_data_sets.md`.

3 Methods

Two different approaches were tested.

1. A neuronal network with an input layer, several hidden layers with dropout and a dense output layer. An effort was made to optimize the param-

eters of the hidden layers and the droupout rates using optuna, a hyperparameter optimization framework. This model was implemented using Keras and Tensorflow. The optimization suggested a network with two units with 128 neurons in one layer (dropout rate 0.158) and 64 neurons in the second layer (dropout rate 0.264).

2. A combination of LGBM (Light Gradient Boosting Machine, a decision tree based algorithm) and SAKT (Self-Attentive model for Knowledge Tracing) [15]. SAKT architecture was similar to that published by Pandey et al. [15], i.e. it consisted of an embedding, a self-attention layer, a feed forward and a prediction layer. This model was implemented using LGB and Torch.

Both methods were taken from solvers available in the competition.

4 Results

Method 1 - Classical feed forward network

Hyperparameter optimization using the optuna framework suggested relatively good results for 2 inner layers with 128 neurons each or with 128 and 64 neurons.
The table below shows the three best neural network architectures found.

Network architecture	Dropout rates	Score
128,128	0,176112,0,213058	0,7670979
128,64	0,150727-0,126969	0,7670460
128,128	0,155383,0,168237	0,7670460
128,128	0,174529,0,219706	0,7669651
128-64	0,162066,0,110946	0,7669593

Figure 1: Results of the optuna optimization of the network architecture

Figure 2 shows the accuracy of predictions using the network selected by the optimization procedure. The best validation accuracy obtained was 0.7755, the overall accuracy score was 0.76766.

Method 2 - Combined LGBM/SAKT approach

With combined LGBM/SAKT approach the best prediction accuracy obtained was 0.884 for the training set and 0.772 for the validation set. Figure 3 shows the most important parameters for performance prediction.

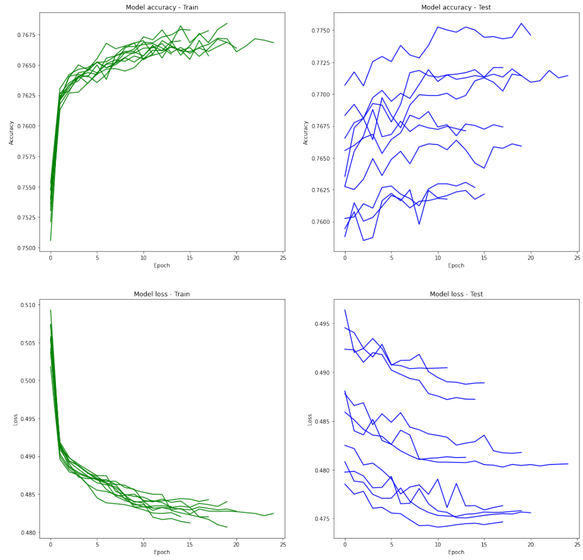

Figure 2: Results obtained with the 128–64 network architecture.

importance.png

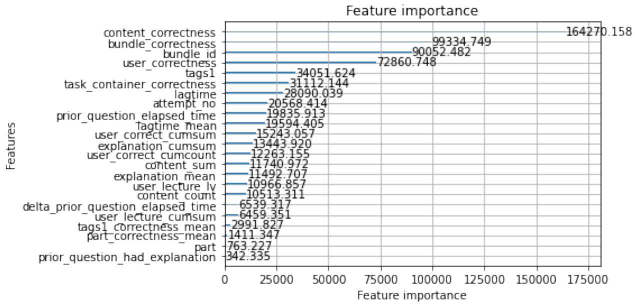

Figure 3: The importance of individual parameters (LGBM/SAKT method)

5 Conclusion

Two distinct methods were tested to solve the problem of Answer Correctness Prediction, a classical feed forward neural network with hyperparameter optimization and a combined LGBM/SAKT approach. The LGBM/SAKT approach may provide slightly better results in performance prediction, however, many parameters of both methods should be further optimized. Prediction accuracy was still below 0.8. So far, only one of the goals of the competition was attempted to be solved. Further effort should be made to fulfill the remaining two aims, as well.

References

[1] Kaggle answer correctness prediction task. online, 2020. [cit. 2020–

11–29] https://www.kaggle.com/c/riiid-test-answer-prediction.

[2] Ghodai Abdelrahman and Qing Wang. Knowledge tracing with sequential key-value memory networks. 2019.

[3] John Anderson, Albert Corbett, Kenneth Koedinger, and Ray Pelletier. Cognitive tutors: Lessons learned. *Journal of the Learning Sciences*, 4:167–207, 04 1995.

[4] Haw-Shiuan Chang, Hwai-Jung Hsu, and Kuan-Ta Chen. Modeling exercise relationships in e-learning: A unified approach. In *EDM*, pages 532–535, 2015.

[5] Youngduck Choi, Youngnam Lee, Dongmin Shin, Junghyun Cho, Seoyon Park, Seewoo Lee, Jineon Baek, Chan Bae, Byungsoo Kim, and Jaewe Heo. Ednet: A large-scale hierarchical dataset in education, 2020.

[6] Albert T Corbett and John R Anderson. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction*, 4(4):253–278, 1994.

[7] Mingyu Feng, Neil Heffernan, and Kenneth Koedinger. Addressing the assessment challenge with an online system that tutors as it assesses. *User Modeling and User-Adapted Interaction*, 19(3):243–266, 2009.

[8] Zhenya Huang, Qi Liu, Chengxiang Zhai, Yu Yin, Enhong Chen, Weibo Gao, and Guoping Hu. Exploring multi-objective exercise recommendations in online education systems. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pages 1261–1270. ACM, 2019.

[9] Zhenya Huang, Yu Yin, Enhong Chen, Hui Xiong, Yu Su, Guoping Hu, et al. Ekt: Exercise-aware knowledge tracing for student performance prediction. *IEEE Transactions on Knowledge and Data Engineering*, 2019.

[10] Kenneth R Koedinger, Ryan SJd Baker, Kyle Cunningham, Alida Skogsholm, Brett Leber, and John Stamper. A data repository for the edm community: The pslc datashop. *Handbook of educational data mining*, 43:43–56, 2010.

[11] Youngnam Lee, Youngduck Choi, Junghyun Cho, Alexander R Fabbri, Hyunbin Loh, Chanyou Hwang, Yongku Lee, Sang-Wook Kim, and Dragomir Radev. Creating a neural pedagogical agent by jointly learning to review and assess. *arXiv preprint arXiv:1906.10910*, 2019.

[12] Qi Liu, Shiwei Tong, Chuanren Liu, Hongke Zhao, Enhong Chen, Haiping Ma, and Shijin Wang. Exploiting cognitive structure for adaptive learning. *arXiv preprint arXiv:1905.12470*, 2019.

[13] Rose Luckin, Wayne Holmes, Mark Griffiths, and Laurie B Forcier. Intelligence unleashed: An argument for ai in education. 2016.

[14] Travis Scott Mandel. *Better Education Through Improved Reinforcement Learning*. PhD thesis, 2017.

[15] Shalini Pandey and George Karypis. A self-attentive model for knowledge tracing. *arXiv preprint arXiv:1907.06837*, 2019.

[16] Zachary A Pardos, Ryan SJD Baker, Maria OCZ San Pedro, Sujith M Gowda, and Supreeth M Gowda. Affective states and state tests: Investigating how affect and engagement during the school year predict end-of-year learning outcomes. *Journal of Learning Analytics*, 1(1):107–128, 2014.

[17] Radek Pelánek. Bayesian knowledge tracing, logistic models, and beyond: an overview of learner modeling techniques. *User Modeling and User-Adapted Interaction*, 27:313–350, 2017.

[18] Chris Piech, Jonathan Bassen, Jonathan Huang, Surya Ganguli, Mehran Sahami, Leonidas J Guibas, and Jascha Sohl-Dickstein. Deep knowledge tracing. In *Advances in neural information processing systems*, pages 505–513, 2015.

[19] Siddharth Reddy, Sergey Levine, and Anca Dragan. Accelerating human learning with deep reinforcement learning. In *NIPS'17 Workshop: Teaching Machines, Robots, and Humans*, 2017.

[20] Doaa Shawky and Ashraf Badawi. A reinforcement learning-based adaptive learning system. In *International Conference on Advanced Machine Learning Technologies and Applications*, pages 221–231. Springer, 2018.

[21] Sugandh Sinha. Using deep reinforcement learning for personalizing review sessions on

e-learning platforms with spaced repetition, 2019.

[22] M. Yudelson. Individualizing bayesian knowledge tracing. are skill parameters more important than student parameters? In *EDM*, 2016.

[23] M. Yudelson, K. Koedinger, and G. Gordon. Individualized bayesian knowledge tracing models. In *AIED*, 2013.

[24] Jiani Zhang, Xingjian Shi, Irwin King, and Dit-Yan Yeung. Dynamic key-value memory networks for knowledge tracing. In *Proceedings of the 26th international conference on World Wide Web*, pages 765–774. International World Wide Web Conferences Steering Committee, 2017.

[25] Guojing Zhou, Hamoon Azizsoltani, Markel Sanz Ausin, Tiffany Barnes, and Min Chi. Hierarchical reinforcement learning for pedagogical policy induction. In *International Conference on Artificial Intelligence in Education*, pages 544–556. Springer, 2019.