Deep learning for knowledge tracing in learning analytics: an overview

Gabriella Casalino[a], Luca Grilli[b], Pierpaolo Limone[c], Domenico Santoro[d] and Daniele Schicchi[e,f]

[a]*University of Bari, Department of Computer Science, Bari, Italy*
[b]*University of Foggia, Department of Economics, Management and Territory, Foggia, Italy*
[c]*University of Foggia, Department of Humanities, Foggia, Italy*
[d]*University of Bari, Department of Economics and Finance, Bari, Italy*
[e]*University of Palermo, Department of Computer Science, Palermo, Italy*
[f]*Dipartimento di Scienze per l'Innovazione Tecnologica, Istituto Euro-Mediterraneo di Scienza e Tecnologia, Palermo, Italy*

Abstract
Learning Analytics (LA) is a recent research branch that refers to methods for measuring, collecting, analyzing, and reporting learners' data, in order to better understand and optimize the processes and the environments. Knowledge Tracing (KT) deals with the modeling of the evolution, during the time, of the students' learning process. Particularly its aim is to predict students' outcomes in order to avoid failures and to support both students and teachers. Recently, KT has been tackled by exploiting Deep Learning (DL) models and generating a new, ongoing, research line that is known as Deep Knowledge Tracing (DKT). This was made possible by the digitalization process that has simplified the gathering of educational data from many different sources such as online learning platforms, intelligent objects, and mainstream IT-based systems for education. DKT predicts the student's performances by using the information embedded in the collected data. Moreover, it has been shown to be able to outperform the state-of-the-art models for KT. In this paper, we briefly describe the most promising DL models, by focusing on their prominent contribution in solving the KT task.

Keywords
Knowledge Tracing, Machine Learning, Deep Learning, Learning Analytics, Educational data, Students skills

1. Introduction

Educational Informatics refers to a research branch where computer science technologies meet educational theories [1]. In recent years, the advent of the digitalization has drastically affected the educational domain, as well as several other fields, and different IT-based solutions for Education have emerged. Some examples are the use of virtual reality [2], social robots [3, 4], video processing [5], text analysis [6], cloud computing [7], Internet of Things (IoT) and wearable technologies [8, 9] , just to mention few. Moreover new learning technologies have exponentially

teleXbe - Technology Enhanced Learning Environments for Blended Education, January 21–22, 2021, Foggia (Italy)
luca.grilli@unifg.it (L. Grilli); daniele.schicchi@unipa.it (D. Schicchi)
0000-0003-0713-2260 (G. Casalino); 0000-0003-0931-2054 (L. Grilli); 0000-0003-3852-4005 (P. Limone); 0000-0001-9505-0673 (D. Santoro); 0000-0003-0154-2736 (D. Schicchi)

grown, such as e-learning, online virtual learning environments (VLE), massive open online course (MOOC), blended learning. If from one hand this has led to new opportunities, such as a big amount of daily generated data obtained from the interaction between the students and the learning technologies, on the other hand it has posed new challenges, such as how to deal with this kind of data.

Learning Analytics is a recent research branch with the aim of understanding and optimizing the students' learning process. The evolving processes that are involved in the learning activity are studied and the relationships between the students' outcomes and their interactions with the virtual platforms are also considered [10].

Data coming from virtual environments are widely used to model students' behavior in order to predict their performances while avoiding failures. Particularly, machine learning techniques have been proved to be effective in analyzing this big amount of data [11], coming from different sources [12], and changing during the time [13].

The insights derived from the educational data are used to support students' learning and better design the courses in order to master the most of the students [14]. Moreover these results are commonly used by decision-makers in order to manage and construct the courses offered at a higher educational level.

The two main theories that are used to model the students' knowledge in order to predict their performance are the *Item Response Theory* (IRT) [15], and the *Knowledge Tracing* (KT) [16]. Both theories describe the students' knowledge in terms of skills that they are supposed to possess in order to successfully perform a task (that is usually formalized in a questionnaire item). This was previously theorized by Bloom in [17] where he discussed what the student is required to know in order to master a subject. He had the two fold aim to define the teaching activities to make the most students able to master a subject (over than 90%) and to determine how individual differences in learners can be related to the learning and teaching process.

Whilst IRT gives a static image of the students' learning, based on Q-matrix that maps student' skills with questionnaire items, on the contrary, KT takes into account how the students' abilities changes over the time [18]. On the other hand, IRT analyzes all the students at the same times whilst KT analyzes students separately. Since the manually mapping, performed by domain experts in Q-matrices, was both a very expensive task as well as subjective [19], thus automatic techniques were proposed [20]. However this students' representation while takes into account differences among students and items, it is not able to model the students' performance over the time. Recent works have applied Deep learning techniques to Knowledge tracing (*Deep Knowledge Tracing* – DKT) in order to model single student's behavior during the time, in terms of hidden skills that are automatically extracted from data [21]. Moreover, these approaches have been shown to be able predict the students' outcomes better than the *Bayesian Knowledge Tracing* approaches (BKT) that were used before [22].

In this paper, we focus on the most promising deep learning models for the KT problem. We qualitatively explain how they work, by avoiding technical details. Indeed the aim is to make the reader aware of the potentiality that deep learning offers for tackling such a problem.
The rest of the paper is structured as follows. Section 2 introduces the reader into Machine-Deep Learning methodologies, then Knowledge Tracing theory is summarized in section 3. An overview of DL algorithms for KT is sketched in section 4. Section 5 concludes the paper and depicts future developments of the research.

2. Machine Learning and Deep Learning

Tackling problems in computer science is done by designing *algorithms*: a finite sequence of computer-implementable instructions which lead to the problem solution. Most of the time, this process is carried out by analyzing the original problem and then by finding a set of rules that must be applied to solve the task. For non-trivial problems, this process can lead to a very long list of complex rules, that makes the computer programs hard to maintain. Moreover, it is not always possible to identify an exhaustive set of rules.

The basic idea of *Machine Learning* (ML) is to go beyond the traditional way of programming, by leaving the machines the task to figure out the algorithms' steps. In order to achieve this task, a huge amount of data, that is representative of a specific problem, is given to the machine in order to *learn* the way for solving it. Technically, data is utilized for tweaking the model's parameters (hereinafter *weights*) which define the system's behavior.

Deep Learning (DL) is a ML sub-field that tackles tasks by exploiting a specific class of models named *artificial neural networks* (ANNs). In the last decade, DL has captured the attention of the entire scientific community since ANN-based systems have shown unbelievable performances while addressing hard problems. DL has become a hot research topic and a big variety of powerful models that can deal with a wide range of problems, have been developed.

Recurrent Neural Networks (RNNs) have been a breakthrough for the analysis of temporal sequences. They are a class of Neural Networks (NN) capable of handling sequential data for predictive tasks. Narrowly, their functioning makes use of a *feedback* connection that links the outcome computed by the network for a single sequence's element to the input of the network itself. Such an approach was shown to be effective in allowing the network to take into account both the nature and the structure of the sequence. Moreover, the RNNs have been utilized for different tasks where the time factor is relevant. For example they have been used to classify documents and sentences [23, 24], to predict trajectories of a vehicle for avoiding accidents [25], or to create an acoustic model for speech recognition [26].

Recent works [27] have shown that the classical RNN architecture has some problems in learning long-term dependencies. The problem was fixed through the design of new artificial neuron structures like the *Long Short Term Memory* (LSTM) one [28]. Furthermore, the introduction of the *attention* concept [29] combined with the enhanced RNNs architecture have significantly improved the performance of DL for temporal sequence-based tasks. Such a concept leverages the idea of discovering the relevance of sequence items, that is elements that are most relevant for fulfilling the prediction task.

Recently, a new class of NNs, named *transformer*, has been presented [30]. Transformer deals with temporal sequences while avoiding the usage of feedback connections but relying extensively on the attention mechanism to find dependencies between input and output. Its performances have been shown to widely overcome the RNN ones and nowadays it represents the state of the art for handling sequence-related problems.

3. Knowledge Tracing

Knowledge Tracing (KT) was firstly proposed by Corbett and Anderson in their ACT Programming Tutor (APT) that was meant to guide the students in Lisp programming tasks [16]. The basic idea behind the model was that a student would be able to master a subject if two conditions are satisfied: the domain knowledge should be organized in a hierarchical structure of skills, and this hierarchy should be proposed during the learning experience, so as students can master low level skills before approaching to the highest level skill. The model assumes that students firstly acquire skill knowledge in a declarative form (from reading, for example), and then domain-specific procedural knowledge is acquired by means of practical tasks resolutions. They defined an *ideal student model* as a set of the rules (skills and sub-skills) that the student should have known in order to acquire the domain knowledge. At each time the probability that the student has learned each rule in the ideal model is evaluated and if needed, a guidance is provided to the student, to successfully perform a given task. Knowledge tracing is used in the APT to implement mastery learning, i.e. assessment is continuously integrated with practice [16].

Thus Knowledge Tracing is a sequence of student's outcome predictions on given questions, while he is still interacting with the virtual learning platform.

Since KT is based on the probability theory, the first algorithms to tackle this problem where Bayesian classifiers (Bayesian Knowledge Tracer – BKT) where the probability was evaluated at each time T [31]. Particularly for each item of a questionnaire, four different scenarios could happen: the student correctly answered the question since it has acquired the necessary knowledge, he failed in answering the question because he does not have the required skills, or he correctly answered because he guessed the response, or on the contrary even if he had the skills to correctly answer to a question, he made a mistake. So, given the history of the student, at each time T, the probability that one of these scenarios would occur is evaluated.

Recently Deep learning based approaches for Knowledge Tracing (DKT) have been proposed [21] and they have been proven to overcome the accuracy of BKTs [22]. Students are modeled separately. Each student is described by the predicted probabilities he will use a skill to solve an exercise. Hidden skills are automatically extracted from data, and at each time T, the model has learnt the historical student's data and it is able to predict the probability to master the next item, and which skills will be involved, together with their probability. In this way it is possible to detect students that need further attention or recommend learning resources based on the skills that have not been acquired yet [32].

Please note that in order to make KT effective, a learning activity to be analyzed should comply with the two conditions Corbett and Anderson have defined. Thus leaning materials and activities, both online or in presence, should be incremental, and they should require knowledge that is supposed to be previously acquired from the student. Thus the learning activities should follow the incremental nature of the student's learning process.

4. Deep Knowledge Tracing Overview

Classical Recurrent Neural Network approaches

RNNs have been widely utilized to tackle the KT problem showing great achievements to predict how the student will perform through the analysis of his past interactions. Technically, such models analyze question-answer pairs $\{q_t, a_t\}$ over time to learn to predict the student's answer at time $t + 1$. The first attempt was carried out by Piech et al. [21] who introduced formally the concept of *Deep Knowledge Tracing*. The authors pointed out to formulate the task as a temporal application since the student's knowledge increases along the time dimension. Their experimentation has investigated both classical and LSTM RNNs to analyze the data representing the student's history in order to trace the acquired knowledge and to predict his/her future performances. Such an approach avoid hard-coded the student competencies which is a demanding task that involves expert annotators.

In this first work, three different sets of data have been separately utilized to *train* models to trace the student knowledge: *Simulated-5*, *Khan Math*, and *Assistments "skill builder"*. The first one simulates the answers of 4.000 virtual students of 50 exercises relative to 5 different concepts. The knowledge of each virtual students is modeled through the Item Response Theory [15], and their skills increase over time. Each exercise regards a single concept and it is marked with a difficulty level. Khan Math is a collection of data gathered from the Khan Academy which includes 1.4 million exercises carried out by 47.495 students of 69 different types. The last one is a set of data that comes from ASSISTments, an Intelligent Tutor System that supports students to tackle problems related to math. A record of actions computed by the student is outputted from the tutor every time the exercise is completed correctly. Such data concerns students in the time range 2009-2010, it is publicly available and represents one of the biggest resource suitable for addressing the KT problem employing ML.

Evaluation of the models' performances has considered a comparison with the Bayesian Knowledge Tracing, a state of the art measure, by computing the *Area Under the Curve Receiving Operating Characteristic Curve* (AUC). AUC is a trusted measure for assessing the quality of ML models and it is often utilized also for making comparisons. The Piech et al. work showed a performance gain up to 25% of the RNNs, attesting promising results in using such class of models.

Subsequently, Xiong et al. [22] have revised the preliminary score presented by Piech et al. [21] uncovering issues that were not considered. First, they found that the ASSISTments 2009-2010 showed the 23.6% of duplicated data that were erroneously utilized for the analysis of the model. The usage of this large amount of identical data does not allow to test the model behavior fairly invalidating the performance measurements. Second, DKT leveraged the advantage of exploiting the *scaffolding* features that competitors have not taken into account. Last, in the ASSISTments 2009-2010 assignments tagged with multiple skills were divided into more chunks of data addressing a specific skill. This means that if an assignment had been tagged with many different skills, the data would have included the same set of actions repeated as many time as the number of skills, including additional information not exploited by competitors advantaging the RNNs functioning.

The authors have considered the issues, and they tested the RNNs on more reliable sets of data.

The final results show an overall better behavior achieved by using RNNs, but the performance gap was reduced.

The innovation introduced by the DKT has led scientists to investigate qualitative differences regard previous models like the BKT. Khajah et al. [33] highlight different components that DKT is capable of considering but BKT is not. They proved that RNNs embed the *recency effect*, that is the model fits the human-beings reasoning processes since it is affected by recent events more than past ones. DTK can *contextualized trial sequences* since its input is the sequence of exercises a student receives in the same order. This allows to infer how the exercise sequences affect the student's learning. Moreover, DKT is capable of discovering *inter-skill similarity*, i.e the degree of relatedness among skills for solving the exercise, and understanding *individual variation in ability* since it can use a student's history achievements to predict the performance for the next exercise.

Enhanced Deep Knowledge Tracing

BKT and DKT fail to identify if a certain concept is mastered by a given student. To solve this problem, Zhang et al. [34] proposed a new model, named Dynamic Key-Value Memory Networks (DKVMN), capable of learning the relationships between concepts in order to output the student's comprehension for each of them. DKVMN is inspired by the *Memory-Augmented Neural Network*[35], a particular NN which exploit an external memory module for enhancing the ability of the model to capture long-term dependencies. DKVMN makes use of two memory modules: the first one is a static matrix including the concepts knowledge (*key*), the second one is a dynamic matrix which contains the student's competencies for each concept (*value*). Such an approach has been compared with the classical DKT model proposed by Piech et al. [21] on four different sets of data. Results have shown better performance achieved by the DKVMN for tracing the student's knowledge, providing a detailed schema concerned the concepts mastered level for each student.

A different way to improve the functioning of RNNs for the KT problem has been proposed by Zhang et al. [36]. They propose to look at the breadth of other features recorded by the computer-based learning platforms and including them in DKT models. Authors experimented the usage of the amount of time students need to respond (*students response time*), the number of attempts to complete the exercise correctly (*attempt number*), and if the first action made by the student was for requesting help (*first action*). Then, they proposed a way to represent such additional information in order to make them available for the analysis through RNNs. Augmenting the number of the considered features have been shown to outperform the original DKT model. In this case, the original DKT structure was subjected to minor changes, underling that a richer set of students' information, describing also the contexts of their actions, is relevant to achieve better results.

Nowadays, *Transformers* models (see sec. 2) are the state of the art for tackling temporal sequences tasks by DL. Tackling the KT has exploited such models in several respects, though it is an emerging research topic. Pandey et al. [37] carried out the first attempt to improve DKT by exploiting self-attention based NNs. Their approach, SAKT, identifies the mastered concepts by analyzing the student's past activities. In addition, it faces better than previous DL-based systems the sparsity data problem (i.e students interact with few concepts generating a small

amount of information).
The SAKT functioning computes weights (i.e attention weights) to assign to the completed exercises during the prediction of the student's performance on a certain exercise. The visualizations of the attention weights for a certain prediction helps to understand which of the completed exercises the network leverages for choosing the outcome of the student, that is what are the past relevant exercises for solving the current exercise. SAKT has been widely tested on real-world datasets, and compared with previous DL models showing on average an improvement of the 4.43% on the AUC.

Transformers do not make of use of the *feedback* connection tracking time by exploiting the *positional encoding* [30]. Recent advancements have led to the development of a novel architecture named *Transformer-XL*. It involves a recurrence mechanism and a new positional encoding scheme which makes it capable of capturing longer-term dependency much better than both RNNs and classical Transformer models.
He et al. [38] exploit the characteristics of the Transformer-XL for overcoming problems related to the analysis of long input exercise sequences which affect negatively the performances of the past DL models. Their system, named KT-XL, has been extensively tested on three real-world data sets, and then compared with past models like DKT, DKVMN and SAKT have been carried out. KT-XL achieves the best performances over all the data sets, showing an averaged improvement of 3.6%.

5. Conclusion

In this paper, the use of deep learning techniques for Learning Analytics has been described. Particularly we focused on the Knowledge Tracing methodology, that is used to predict the probability a student will successfully perform an exercise, given the history of his previously performed exercises. KT is able to model the student's behavior over the time, and its results can be used as feedback by both students and teachers. Indeed it could be used to warn the teacher in case the student has not mastered the skills that are required by a given subject. Thus almost real-time interventions could be adopted to support the student's learning process. Moreover the teacher could perhaps modify the design of his course in order to better clarify some concepts for which the most of the students shown difficulties. The Knowledge tracing theory is effective for both in presence and online learning.

Particularly, blended learning models, that combine on-line with face-to-face activities, could benefit from this assessment evaluation approach. Indeed, if from one hand the use of technology mediated activities enhances the learning experience, from the other hand, when the face-to-face settings is left, grasping students' engagement is more difficult [39]. Thus, automatic techniques that are able to monitor students' learning, while interacting with online activities, could be a useful aid to teachers for sensing their classroom engagement and for detecting changes in learning behavior of single students.

Nowadays, DL have shown to be the most promising class of models for tackling the KT problem. Although, the mixture between DL and KT is an emerging research topic, their performances have overcome the ones of the past systems which exploit different models (e.g Bayesian algorithms). Simple RNN have been shown to be effective for tracing the student's knowledge by examining his/her history, but the topic's relevance have led to test and tune DL alternatives such us Transformers that have become the state-of-the-art methodology for this task.

Future work will be addressed to systematically review Deep Knowledge Tracing literature, by highlighting strengths and weaknesses of each method, the available educational data, common evaluation measures, and further directions to address.

Finally, since the stakeholders involved in the Educational domain are not necessarily technicians, explainable mechanisms should be considered in order to improve the interpretation of the results achieved by the learning algorithms [40, 41].

Acknowledgements

Gabriella Casalino acknowledges funding from the Italian Ministry of Education, University and Research through the European PON project AIM (Attraction and International Mobility), nr. 1852414, activity 2, line 1. Gabriella Casalino and Daniele Schicchi are members of the INdAM GNCS research group.

References

[1] G. Casalino, G. Castellano, A. Mannavola, G. Vessio, Educational stream data analysis: A case study, in: 2020 IEEE 20th Mediterranean Electrotechnical Conference (MELECON), 2020, pp. 232–237.

[2] M. Arrigo, F. Cappello, F. Di Paola, M. Farella, G. Lo Bosco, D. Saguto, F. Sannasardo, HeARt mobile learning, in: 10th Annual International Conference on Education and New Learning Technologies (EDULEARN 18), IATED Academy, 2018, pp. 10899–10905.

[3] A. Vitanza, P. Rossetti, F. Mondada, V. Trianni, Robot swarms as an educational tool: The thymio's way, International Journal of Advanced Robotic Systems 16 (2019) 1729881418825186.

[4] D. Schicchi, G. Pilato, A social humanoid robot as a playfellow for vocabulary enhancement, in: 2018 Second IEEE International Conference on Robotic Computing (IRC), IEEE, 2018, pp. 205–208.

[5] B. De Carolis, F. D'Errico, N. Macchiarulo, G. Palestra, Engaged faces: Measuring and monitoring student engagement from face and gaze behavior, in: IEEE/WIC/ACM International Conference on Web Intelligence-Companion Volume, 2019, pp. 80–85.

[6] T. Aljohani, F. D. Pereira, A. I. Cristea, E. Oliveira, Prediction of users' professional profile in moocs only by utilising learners' written texts, in: International Conference on Intelligent Tutoring Systems, Springer, 2020, pp. 163–173.

[7] M. Scalera, E. Gentile, P. Plantamura, G. Dimauro, A systematic mapping study in cloud for educational innovation, Applied Sciences 10 (2020) 4531.

[8] W. Guest, F. Wild, D. Di Mitri, R. Klemke, J. Karjalainen, K. Helin, Architecture and design patterns for distributed, scalable augmented reality and wearable technology systems, in: IEEE International Conference on Engineering, Technology and Education, IEEE TALE 2019: Creative & Innovative Education to Enhance the Quality of Life, 2019.

[9] G. Tripathi, M. A. Ahad, Iot in education: an integration of educator community to promote holistic teaching and learning, in: Soft Computing in Data Analytics, Springer, 2019, pp. 675–683.
[10] M. Brown, Learning analytics: The coming third wave, EDUCAUSE Learning (2011).
[11] C. Romero, S. Ventura, Educational data mining and learning analytics: An updated survey, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 10 (2020) e1355.
[12] D. Di Mitri, J. Schneider, M. Specht, H. Drachsler, Multimodal pipeline: A generic approach for handling multimodal data for supporting learning, in: First workshop on AI-based Multimodal Analytics for Understanding Human Learning in Real-world Educational Contexts, 2019.
[13] G. Casalino, G. Castellano, G. Vessio, Exploiting time in adaptive learning from educational data, in: International Workshop on Higher Education Learning Methodologies and Technologies Online, Springer, 2021.
[14] G. Castellano, A. Fanelli, T. Roselli, Mining categories of learners by a competitive neural network, in: IJCNN'01. International Joint Conference on Neural Networks. Proceedings (Cat. No. 01CH37222), volume 2, IEEE, 2001, pp. 945–950.
[15] F. Lord, A theory of test scores., Psychometric monographs (1952).
[16] A. T. Corbett, J. R. Anderson, Knowledge tracing: Modeling the acquisition of procedural knowledge, User modeling and user-adapted interaction 4 (1994) 253–278.
[17] B. S. Bloom, Learning for mastery. instruction and curriculum. regional education laboratory for the carolinas and virginia, topical papers and reprints, number 1., Evaluation comment 1 (1968) n2.
[18] M. M. Khajah, Y. Huang, J. P. González-Brenes, M. C. Mozer, P. Brusilovsky, Integrating knowledge tracing and item response theory: A tale of two frameworks, in: CEUR Workshop proceedings, volume 1181, University of Pittsburgh, 2014, pp. 7–15.
[19] K. K. Tatsuoka, Rule space: An approach for dealing with misconceptions based on item response theory, Journal of educational measurement (1983) 345–354.
[20] G. Casalino, C. Castiello, N. Del Buono, F. Esposito, C. Mencar, Q-matrix extraction from real response data using nonnegative matrix factorizations, in: International Conference on Computational Science and Its Applications, Springer, 2017, pp. 203–216.
[21] C. Piech, J. Bassen, J. Huang, S. Ganguli, M. Sahami, L. J. Guibas, J. Sohl-Dickstein, Deep knowledge tracing, Advances in Neural Information Processing Systems (2015) 505––513.
[22] X. Xiong, S. Zhao, E. G. V. Inwegen, J. E. Beck, Going deeper with deep knowledge tracing, Proceedings of the 9th International Conference on Educational Data Mining (2016) 545––550.
[23] G. L. Bosco, G. Pilato, D. Schicchi, A sentence based system for measuring syntax complexity using a recurrent deep neural network., in: NL4AI@ AI* IA, 2018, pp. 95–101.
[24] G. L. Bosco, G. Pilato, D. Schicchi, A neural network model for the evaluation of text complexity in italian language: a representation point of view, Procedia computer science 145 (2018) 464–470.
[25] B. Kim, C. M. Kang, J. Kim, S. H. Lee, C. C. Chung, J. W. Choi, Probabilistic vehicle trajectory prediction over occupancy grid map via recurrent neural network, in: 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), 2017, pp. 399–404. doi:10.1109/ITSC.2017.8317943.

[26] H. Sak, A. Senior, K. Rao, F. Beaufays, Fast and accurate recurrent neural network acoustic models for speech recognition, 2015. arXiv:1507.06947.
[27] Y. Bengio, P. Simard, P. Frasconi, Learning long-term dependencies with gradient descent is difficult, IEEE Transactions on Neural Networks 5 (1994) 157–166. doi:10.1109/72.279181.
[28] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural computation 9 (1997) 1735–1780.
[29] D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, 2014. arXiv:1409.0473.
[30] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Advances in neural information processing systems 30 (2017) 5998–6008.
[31] B. van De Sande, Properties of the bayesian knowledge tracing model., Journal of Educational Data Mining 5 (2013) 1–10.
[32] J.-J. Vie, H. Kashima, Knowledge tracing machines: Factorization machines for knowledge tracing, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 33, 2019, pp. 750–757.
[33] M. Khajah, R. V. Lindsey, M. C. Mozer, How deep is knowledge tracing?, arXiv:1604.02416v2 (2016).
[34] J. Zhang, X. Shi, I. King, D.-T. Yeung, Dynamic key-value memory networks for knowledge tracing, WWW 17: Proceedings of the 26th International Conference on World Wide Web (2017) 765––774.
[35] A. Graves, G. Wayne, M. Reynolds, T. Harley, I. Danihelka, A. Grabska-Barwińska, S. G. Colmenarejo, E. Grefenstette, T. Ramalho, J. Agapiou, et al., Hybrid computing using a neural network with dynamic external memory, Nature 538 (2016) 471–476.
[36] L. Zhang, X. Xiong, S. Zhao, A. Botelho, N. T. Heffernan, Incorporating rich features into deep knowledge tracing, LS '17: Proceedings of the Fourth (2017) ACM Conference on Learning (2017) 169––172.
[37] S. Pandey, G. Karypis, A self-attentive model for knowledge tracing, arXiv preprint arXiv:1907.06837 (2019).
[38] Y. He, X. Hu, Z. Zu, G. Sun, Kt-xl: A knowledge tracing model for predicting learning performance based on transformer-xl, ACM TURC'20: Proceedings of the ACM Turing Celebration Conference - China (2020) 175–179. doi:10.1145/3393527.3393557.
[39] R. A. Rasheed, A. Kamsin, N. A. Abdullah, Challenges in the online component of blended learning: A systematic review, Computers & Education 144 (2020) 103701.
[40] J. M. Alonso, G. Casalino, Explainable artificial intelligence for human-centric data analysis in virtual learning environments, in: International Workshop on Higher Education Learning Methodologies and Technologies Online, Springer, 2019, pp. 125–138.
[41] C.-K. Yeung, Deep-irt: Make deep learning based knowledge tracing explainable using item response theory, arXiv preprint arXiv:1904.11738 (2019).