IS DEEP KNOWLEDGE TRACING USING RECURRENT NEURAL NETWORKS EFFECTIVE AT IMPROVING ONLINE EDUCATIONAL PLATFORMS?

BIKOV, KIRIL
Computing Science, College of Science and Engineering

ABSTRACT

The extensive usage of online learning platforms like Moodle among university students creates a vast source of users' behaviour data. Analysing this data with a machine learning model called Deep Knowledge Tracing (DKT) and accordingly adjusting the content and style of online materials can be beneficial for both students and educators.

DKT can detect the pattern of learning actions by analysing data from past student actions. The model can find the most effective way a student learns- for example by practising more exercises or reading additional literature. Since DKT is closely tailored to predict studying behaviour, the model can be incorporated into online educational platforms that automatically adjust the content and materials depending on the needs of the user. The benefit is that a large group of students with different learning methods can make the most out of an online course due to the flexible content generated.

From a lecturer's viewpoint, DKT using Recurrent Neural Networks (an algorithm for learning contextualised data by repeatedly extracting hidden information) can produce data on where students have certain difficulties with a course. By exploring the platform usage, data models can discover the level of user's confusion. The analysed data can be used to inform educators on students' difficulties and allow them to provide just-in-time support. Having up-to-date information on students' struggles allows lecturers to adjust the content of a course, decreasing dropout rates.

This paper explores the effectiveness of DKT for educational platforms. Consequently, the benefits, drawbacks, and the practical applications for all participants in the learning process are presented in the form of research evidence.

INTRODUCTION

In recent years, there has been a major shift in the way education is performed - the role of high school classrooms and university lecture halls has greatly been reduced, as studying transfers to online educational platforms. Studies reveal that universities worldwide are moving more and more towards online learning, where student accessibility and motivation play an integral part for successful functioning of the online educational resources provided (Dr. Wahab Ali, 2020). With the emergence of a global pandemic, all educational institutions, from schools to universities, had to transform their patterns of teaching from classroom-based to online almost overnight. With the global pandemic virtually forcing the integration of online educational platforms, many educators have stated that they will incorporate aspects of virtual learning in their methods for teaching and learning. We have entered a new era of education where online and virtual studying is replacing face-to-face teaching and learning activities (Gauhar Afshan, Aliya Ahmed, 2020). For this reason, it is expected that online educational platforms will continue to play a vital role in students' life, so enhancing such platforms and making maximum and efficient use of them is imperative for better student performance.

The rapid change towards online education because of the COVID-19 pandemic has also had an effect on student engagement. Since students have different learning practices and strengths, not all of them can benefit equally from online educational platforms - studies show that the student drop-out rate in online courses was higher than that in face-to-face learning (De la Fuente et al., 2021; Delgado, 2021). Those results are directly affected by students' engagement, as research evidence concludes that students reported less exposure to effective teaching when using online educational platforms (Dumbford, Miller, 2018). Further studies have found that the longer students engage in learning activities, the better their academic performance (Bravo-Agapito, 2021; Yokoyama, 2019). For this reason, it is crucial that educators can predict and understand students' engagement with online educational platforms. The integration of online educational platforms as a leading way of studying is a vital process for students - improving those platforms to suit the needs and studying habits of students is essential for allowing successful and long-term engagement.

The mass usage of online educational also produces big data which can be reviewed. The term 'big data' refers to a collection of data that is so large in size and complexity that no regular management tools can process it efficiently. In the case of online educational platforms, big data consists of all the students' behaviour when interacting with these platforms. This could, for example, be interaction data (e.g., number of clicks and time between each click), time taken to complete assignments, participation in online forums, or even data on students' demographic and their grades. All these various types of knowledge of the students participating in online educations is contained within the term big data to produce vital information about both the learner's actions and the context in which those actions occur. Such big data can be explored dynamically as it is produced, to extract precise knowledge on the students learning patterns - the knowledge obtained can be directly integrated into the online educational platforms to support students by presenting content in a more efficient manner, closer to the way in which students inherently learn faster and most efficiently.

To utilise the potential of such big data, machine learning algorithms can be applied to analyse the data and produce useful results. These algorithms can be viewed as mathematical model mapping methods used to learn or uncover underlying patterns embedded in the data. In this paper one specific algorithm is explored, which is in the field of deep learning- a type of machine learning based on artificial neural networks in which multiple layers of processing are used to extract progressively high-level features from data. This algorithm is deep knowledge tracing (DKT) using recurrent neural networks (RNN) as a deep learning algorithm to be applied to big data produced from educational platforms. DKT may be able to improve students learning and potentially help educators gain insights into their students' performance is observed

HISTORY OF KNOWLEDGE TRACING(KT) AND CURRENT STATE OF DKT

Knowledge tracing (KT) is an essential task in computer-aided systems, which aims at evaluating students' knowledge over time based on their learning history (Yunfei Liu et al., 2020). The objective of KT is to predict whether students can correctly answer questions based on all their previous responses. In KT, a machine effectively models the knowledge of a student as they interact with studying materials and coursework. KT tries to predict students' ability and how capable they are of correctly answering any future question by observing all their past actions and responses. For example, if a successful student has managed to achieve an impressive score on their coursework so far, then a good implementation of KT should show a very high percentage probability that the student will also perform well on their next assignment. The evolution of KT is long and dates from 1995, when the first such model was proposed, namely Bayesian knowledge tracing (BKT). BKT determines student knowledge using a Hidden Markov Model (HMM) to estimate a set of parameters for each unique skill contained within the data (Christian Fischer et al., 2020). HMM is a model that allow us to predict the probability for a sequence of unknown (hidden) variables from a set of observed variables. An example of an HMM is predicting the weather (hidden variable) based on the type of clothes that someone is wearing (observed). HMM is integrated into BKT so that the chosen parameters describe qualities of the skill being learned, such as how likely students are to guess and what the level of their knowledge is.

Since the model is relatively well established, it has been tested with various data sets and most studies find it to achieve good accuracy results in the range 60-82% AUC (Pu et al., 2020). AUC means Area Under the Curve and it is a measure of the accuracy of the tests by calculating the ability of the model to distinguish between positive (correct) and negative (incorrect) prediction results. The higher the AUC, the better the performance of the model. Research tests (Pu et al., 2020) have been performed on three data sets to inspect the AUC of BKT. The results from table 1 show that the model achieves AUC in the range 62-82% (82% AUC with a smaller dataset and a worse performance on larger datasets with 62.8% and 74.4% AUC):

While BKT uses a Hidden Markov Model to infer student knowledge, Performance Factor Analysis (PFA) uses logistic regression to estimate three parameters for each unique skill within the given data (Pavlik et al., 2009). Logistic regression is a statistical analysis method to predict a binary outcome, such as yes or no, based on prior observations of a data set. A logistic regression model predicts a dependent data variable by analysing the relationship between one or more existing independent variables. For example, a logistic regression could be used to predict whether a political candidate will win or lose an election or whether a high school student will be admitted to a particular college. Compared with BKT, PFA parameters provide less information on the initial knowledge of learners of a given skill, because the binary nature of the model imposes limitations. Moreover, the features used in PFA are relatively simple and they cannot provide a deep insight into students' knowledge (Yeung and Yeung, 2018). However, PFA parameters provide information on the relative difficulty of skills and the relative amount of learning associated with correct and incorrect answers. PFA is still in active improvement and further research is needed to justify its usage.

Table 1. ***AUC results of BKT on three datasets (Pu et al., 2020)***

Datasets	Interactions	Students	Items	Skills	BKT(AUC)
Assistments	943K	1709	4117	102	0.628
STAT F2011	190K	333	1224	81	0.821
KDD 2010	4420K	3287	1379	899	0.744

The most recent model is DKT, first proposed in 2015. DKT uses Recurrent Neural Networks (RNN) to model skill knowledge, producing a vector of the probability of knowledge level. DKT is complex, flexible, and it can discover inter-skill similarities and exercise prerequisites without requiring any specific domain knowledge: like knowing if the course is Spanish or Mathematics, and whether users are university or school students (Sapountzi et al., 2019). It also allows for differences in learning ability of the students by conditioning on the average accuracy for recent learner's performance. DKT is new and still not that well researched, however, it appears to be very successful due to its flexibility in capturing statistical regularities directly present in the inputs and outputs. Compared with the other approaches, DKT is generally more effective at predicting student correctness during learning (Khajah, Lindsay and Mozer, 2016; Yeung and Yeung, 2018), but it has not been used extensively in the real world due to limitations around interpretability and stability of estimates (Yeung and Yeung, 2018). Table 2 shows AUC of BKT and DKT on three data sets produced from online educational platforms (Khajah, Lindsay and Mozer, 2016).

The results from table 2 show that DKT outperforms BKT on every dataset. On the first dataset, DKT has an advantage of 12% and 14% on the second. The difference in model performance on the third data set is smaller (4%), showing that the base DKT also has room for improvement. Nevertheless, the potential of DKT to outperform all previously used KT implementations, often with significant difference, make it a very interesting, new model to explore.

Table 2. ***Performance comparison between BKT and DKT using AUC results (Khajah, Lindsay and Mozer, 2016)***

Dataset	BKT(AUC)	DKT(AUC)
Assistments	0.73	0.85
Synthetic	0.62	0.76
Statics	0.73	0.77

DEEP KNOWLEDGE TRACING DEFINITION AND HOW IT WORKS

Knowledge Tracing can be mathematically formalised as follows: given a student's past interactions Xt = (x1, x2, ..., xT) up to time t on a particular learning task, it predicts some aspects of their next interaction xt+1. In practical scenarios, one such aspect is often how successful the student is at answering coursework questions based on their previous responses and interactions with an online educational platform. To clarify the formula, x1 is a vector of data obtained in a specific period of time that contains information on the specific question which a student has been asked and whether the student has answered it correctly. Xt contains all such previous vectors of data based on the performance of the student so far. Using all this data, the task of Knowledge Tracing is to predict whether a student's next interaction xt+1 with a question will be successful or not. In this case, we specifically use the word 'task', because KT by itself only defines what data is available (all past student responses) and what must be predicted (whether the next response will be correct), but it does not provide an actual implementation.

DKT is an implementation for solving the KT task using recurrent neural networks (RNN) as its backbone. A RNN (Zachary C. Lipton, John Berkowitz, and Charles Elkan. 2015) aims to map the given input sequence (x1, x2,...,xT) to an output sequence (y1,y2,...,yT), visualized in figure 1. During this mapping, the input undergoes a series of transformations via a hidden layer, which captures useful information (that is hard to process by human-engineers), and forms a sequence of hidden states (h1, h2, . . . , hT). This transformation can be stated mathematically as follows:

$$\boldsymbol{h}_t = tanh(\boldsymbol{W}_{hx}\boldsymbol{x}_t + \boldsymbol{W}_{hh}\boldsymbol{h}_{t-1} + \boldsymbol{b}_h)$$

$$\boldsymbol{y}_t = \sigma(\boldsymbol{W}_{hy}\boldsymbol{h}_t + \boldsymbol{b}_y)$$

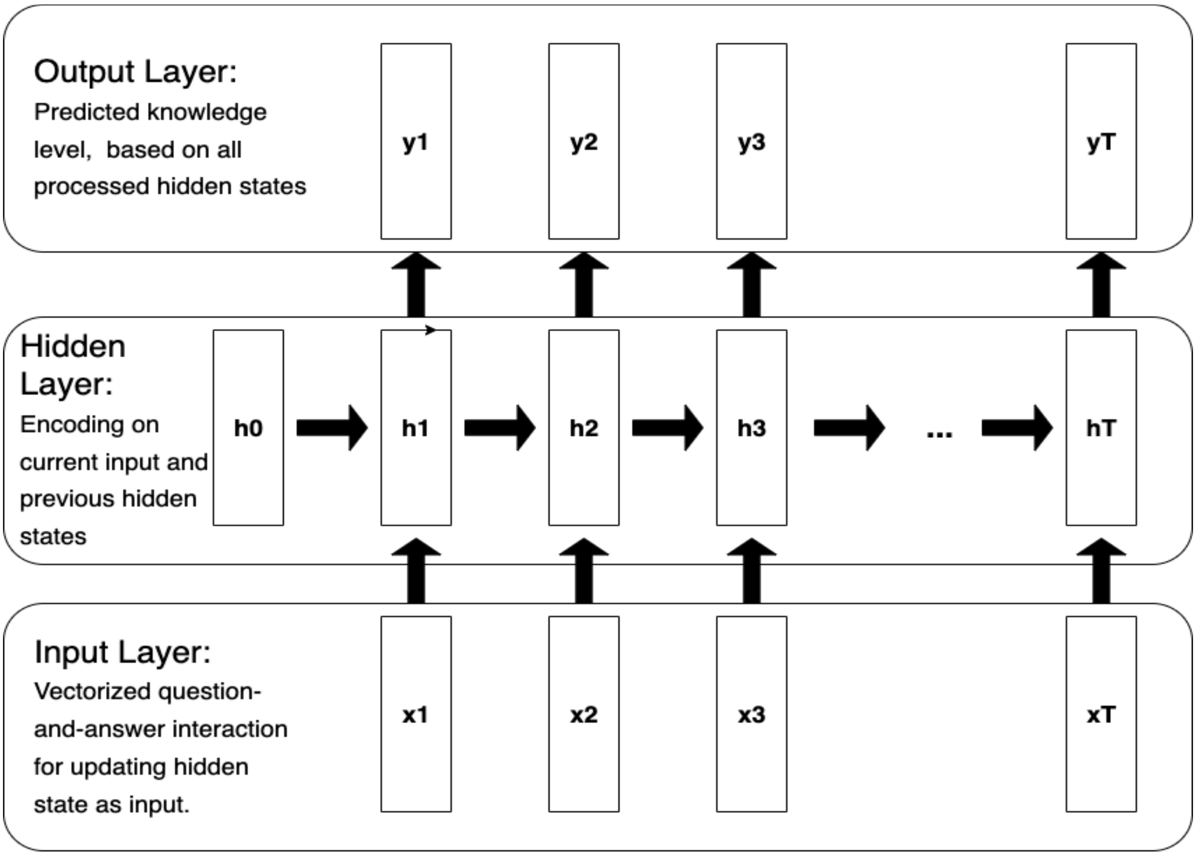

Figure 1. ***A RNN architectural implementation of DKT with input, hidden, and output layer (Yeung and Yeung, 2018)***

To explain the model in a more practical way, the first step starts with an input of previous students' responses (which in the above diagram is called input layer). The input of a student's previous knowledge was already defined in the KT task discussed earlier. Next, this input undergoes various transformation steps that effectively learn new bits of 'hidden' information each time by extracting new context using complex mathematics. Each of the inputs student response x is transformed into information bits containing 'hidden' context h multiple times. On every mapping the model learns new 'hidden' information by also reusing the already obtained context, as visualized below. Finally, the model produces output data which contains all the learned context inside the 'hidden layer':

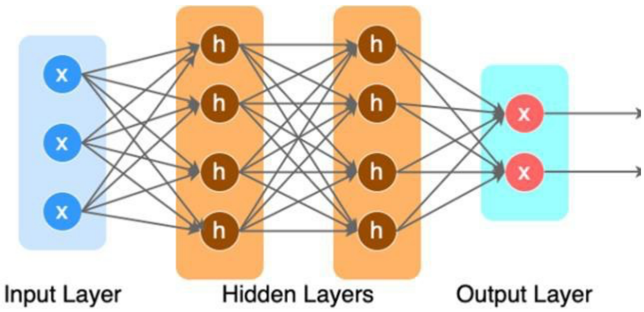

Figure 2. ***Visualization of the relation between layers in RNN***

The hidden layer is what makes the model different and advanced; it allows DKT to obtain data on the students' abilities and knowledge in a specific context. The model learns from previously discovered information inside the hidden layer, which contains multiple mappings as in figure 2. The recurrent neural networks model used by DKT is very closely related to how humans think and process information. It would be much easier for someone to predict whether a student will perform well in the future, if they were given detailed context, such as how confused the student has been feeling lately and whether the student spends more time than their peers to achieve similar results. The hidden layer presented above explores such context and applies it to its final prediction results, making it very advanced and accurate.

DKT can further be extended to use long short-term memory (LSTM). Briefly explained, LSTM is a special kind of RNN which is capable of learning long-term dependencies by remembering past context information for long periods. In regular RNN, each mapping inside the hidden layer performs a single transformation at once, whereas in LSTM a single mapping can be communicating with multiple previous and future mappings to extract additional 'hidden' information. LSTM is an attempt to imitate the human memory system (Christoph Olah, 2015) when extracting context from data, as humans often process current information based on recent discoveries and thoughts.

This paper mainly focuses on DKT using Recurrent Neural Networks. Where only DKT is mentioned, it is assumed that the paper discusses DKT in its RNN implementation.

DKT ADVANTAGES

In many of the experiments conducted on the accuracy of DKT, it demonstrates remarkable performance advantages over its competition, mainly BKT. DKT has the capacity to encode learning context that is outside the scope of BKT. Knowing the specific context of the student's past performance alongside raw data, such as numbers, is what allows DKT to discover structure and dependencies that BKT misses (Khajah, Lindsay, Mozer, 2016).

One area where such DKT context analysis is needed is human behaviour, which is strongly driven by recent past events. For example, when individuals perform a choice task repeatedly, which requires decision making such as answering a quiz question, response time can be predicted by looking into the context of how fast recent task have been performed (H. Ishwaran and L. F. James, 2003). Such recency effects are strongly present in student performance since students must often perform similar tasks repeatedly. For example, when a student is doing a quiz, how fast their previous responses have been often determines the response time for the current question the student is answering. Recurrent neural networks are very effective at discovering and using such human behavioural links. Consequently, DKT is well suited to exploiting recent performance in making predictions.

Furthermore, DKT provides benefits when predicting whether students can successfully complete coursework exercises which require a combination of skills. That is because DKT is fed the entire sequence of exercises and their order is preserved. For this reason, the model can potentially infer the effect of exercise order on learning. DKT also has the capacity to encode inter-skill similarity. If each hidden unit represents student knowledge level for a particular skill, then the degree of overlap can be detected inside the hidden layer of RNN. This feature allows skills to be interpreted also in terms of their degree of relatedness. For example, knowing that a student is good at both math and statistics relates to the probability that the student will also be performing well at computer science coursework. Similarly, DKT can discover interdependence between exercises in the same manner as it discovers interdependence between skills.

Another strong advantage of DKT is its ability to decipher individual variation in ability. Since students vary in ability, individual variation is vital and can potentially be used in a predictive manner. A student’s accuracy in early trials in a sequence might predict accuracy in later trials, regardless of the skills required to solve exercises. DKT provides extensive flexibility in the input features and an ability to identify latent categories (Mongkhonvanit, Kanopka, Lang, 2019), allowing it to correctly interpret individual variations.

DKT can further be improved by incorporating Long Short-Term Memory (LSTM), defined in the previous section. Table 3 shows AUC results of BKT and DKT (implementation with LSTM) for three datasets.

Analysing the results, the introduction of LSTM allows DKT to greatly outperform BKT. In the first dataset, the difference is an impressive 28%, and in the second and third datasets - 17% and 19% correspondingly. The results show that DKT with LSTM can bring significant increase in accuracy.

DKT can also be extended to use transformers, which integrate the ability to learn long-range dependencies (Pu et al., 2020). A transformer is an additional model added to the already used RNN which weighs the significance of each part of the input data. The inclusion of transformers allows the model to learn how to differentiate the importance of frequent and rare questions on students' knowledge level. Research (Pu et al., 2020) shows that DKT with transformer can increase overall model accuracy compared to BKT.

Table 3. ***AUC results of BKT and DKT with an LSTM implementation***

Datasets	Students	Exercise Tags	Answers	BKT(AUC)	DKT(AUC)
Simulated-5	4000	50	200K	0.54	0.82
Khan Math	47495	69	1435K	0.68	0.85
Assisstments	15931	124	526K	0.67	0.86

Table 4. ***AUC Results of BKT and DKT with a transformer (Pu et al., 2020)***

Datasets	Interactions	Students	Items	Skills	BKT(AUC)	DKT Transformers (AUC)
Assistments	943K	1709	4117	102	0.63	0.81
STAT F2011	190K	333	1224	81	0.82	0.95
KDD 2010	4420K	3287	1379	899	0.74	0.79

Table 4 presents prediction accuracy results of BKT and DKT with Transformers on three datasets. DKT achieves 18%, 13% and 5% better results than BKT. Especially impressive is the result of DKT on the second dataset where it manages to obtain a result of 0.95 AUC. The ability of DKT with transformers to learn long-range dependencies can clearly improve prediction results on students' data.

Overall, DKT has important advantages over previous methods because it does not require the explicit encoding of human domain knowledge and can capture more complex representations of student abilities. Many experiments with the model show that it achieves better accuracy than previous models such as BKT and PFA. The high accuracy allows the model to correctly predict the best sequence of learning items to present to students, which could potentially be a very powerful tool for improving online educational platforms.

DKT DRAWBACKS AND THEIR POTENTIAL SOLUTIONS

Although DKT shows many advantages over previous models, there are also some drawbacks that can make the model ineffective. In practical scenarios, DKT often suffers from data scarcity, since some educational institutions have fewer students producing a lower amount of user data (Wu et al., 2021). The data that it produces are also of a different quality because students in different institutions have diverse courses and levels of preparation, which results in unbalanced learning records. Consequently, it is necessary to evaluate the learning data quality before applying it to the model. Sometimes the data are not comparable - it is difficult to compare the knowledge level of students with different learning processes from different educational backgrounds.

A potential solution to these problems is to use an altered framework of the model called Federated DKT (FDKT). In this framework, each client takes charge of training a distributed DKT model (each implementation is independent and separate) and evaluating data quality by leveraging its own local data, while a centre server is responsible for aggregating models and updating the parameters for all the clients (Wu et al., 2021). In this definition, the meaning of client is identical to a single educational institution using DKT in its own online platform. The separate results from different clients(institutions) are then combined, hence the name federated. By having this separation of clients, data quality is evaluated by incorporating different education measurement theories. Another effect of this federation of data is that specific local context can be applied to the model. Tests on three data sets (Wu et al., 2021) show that using Federated DKT outperforms BKT by an average of 20% AUC and base DKT by 10%. These experiments on real-world datasets demonstrate that FDKT brings additional effectiveness to dealing with scarce data of different quality.

Another major limitation of DKT is that the difficulty of the questions which students are asked when using online educational platforms is not always correctly taken into consideration. Questions requiring the same skill may have different difficulties, and thus skill-level prediction cannot always accurately reflect the knowledge level of a student for specific questions. Although it is quite necessary to solve DKT by understanding the difficulty of specific questions, there comes a major issue that the interactions between students and questions are extremely sparse, meaning that there many gaps in the data with missing or not complete information, which can lead to failures in prediction accuracy. The sparseness in interactions between students and questions is due to that fact that a big number of students are answering the same questions at the same time, making it difficult to additionally integrate question difficulty into the prediction model.

To overcome the issue of skill-level prediction based on questions with specific context and tackle sparsity, the underlying information among these questions needs to be extracted in an optimal way. This can be achieved by using an approach called Pre-training Embeddings via Bipartite Graph

(PEBG), which learns a low-dimensional embedding (a map with elements determined by a few of their most important attributes) for each question with all the useful side information (Liu et al., 2020). Embeddings is a method of extracting features out of text so that we can input those features into a machine learning model to work with text data such as questions in online platforms. The term 'pre-training' means that we first extract specific features and how closely they relate (for example words like 'mom' and 'dad' should be closer together than the words 'mom' and 'ketchup'), and then feed this context into DKT. A bipartite graph is mathematical way to structure this feature 'relatedness' into vertices which represent features connected by edges (represent how close features are). To be specific, PEBG includes question difficulties together with three kinds of relations: explicit question-skill relations, implicit question similarity and skill similarity. An example of skill similarity could be the words 'creative' and 'imaginative' being closely related, explicit questions which contain similar key words and are directly related, and implicit ones have different context but can still refer to similar skills. These relations are chosen to maximize the described embedding mechanism. In this way, the learned question embeddings will preserve question difficulty information and the relations among questions and skills. Experiments with question embeddings by PEBG incorporated into DKT performed on three real-world datasets show that using embeddings improves model accuracy by up to 8% (Liu et al., 2020).

There exists another similar limitation of DKT related to questions which students are asked on online educational platforms. All questions nested under a particular skill are treated as equivalent observations of a learner's ability. However, this is an inaccurate assumption in real-world educational scenarios (Sonkar et al., 2020). One of the key assumptions underpinning DKT is that all questions nested under a particular skill are equivalent. This assumption, however, is generally unrealistic in real-world educational datasets. First, a mapping of questions to skills is not always available and obtaining such a mapping requires the intervention of subject matter experts, which is both costly and time-consuming. Second, questions in real-world educational datasets are never exactly equivalent, but rather exhibit significant variations in difficulty and discrimination (Embretson and Reise, 2013). In other words, different questions convey differing levels of information about a particular learner's level of the underlying skill, and methods for modelling a learner's acquisition of skills over time should take such information into account.

Simply substituting questions for skills in a traditional DKT model is insufficient to accomplish the described goal. To achieve understanding of the level of information inside the questions on online educational platforms, a modified model called question-level knowledge tracing (qDKT) can be used (Sonkar et al., 2020). qDKT utilizes a regulariser to incorporate question similarity information. Regularisation is a technique whereby using mathematics (such as variance in success probabilities) in a machine learning model such as DKT is adjusted to minimise any false differences inside the produced data (called loss function). By doing so, regularisation can prevent false assumption in DKT, which can arise due to the dataset having too many questions, by simplifying how DKT calculates relatedness and similarity between questions. qDKT uses a Laplacian matrix (rows and columns of numbers that represent similarity) to calculate the relatedness of questions in the regularisation with the goal of achieving state-of-the-art prediction accuracy results. Unlike base DKT, question-level DKT does not assume that each question must be associated with exactly one skill. The regulariser used by qDKT assumes that success probabilities of multiple questions associated with the same skill should not be significantly different for a given learner.

For even better accuracy performance of the model, word embeddings for model initialization can be used from algorithms like word2vec, fastText and GloVe. These algorithms embed words into a high dimensional space (a map where the position of an element is determined by many attributes) such that words that have close semantic relationships will be embedded near one another, while words with low semantic similarity will be embedded further apart (Goldberg and Levy, 2014). For the case of qDKT, fastText embedding can be used, as it considers individual characters in a word when computing the final embeddings. By doings this, fastText recognises that the words "love", "loved", "lovely", and "lovable" are all related and embed them accordingly.

Table 5. ***AUC Results of qDKT with Laplacian regulariser and fastText word embedding (Sonkar et al., 2020)***

Dataset	DKT	Base qDKT	Base qDKT w/ Laplacian regularizer	Base qDKT w/ fastText	Base qDKT w/ fastText and regularizer
ASSISTments 2009	0.740 ± 0.002	0.678 ± 0.004	0.738 ± 0.003	0.740 ± 0.004	$\mathbf{0.762 \pm 0.005}$
ASSISTments 2017	0.721 ± 0.002	0.742 ± 0.003	0.753 ± 0.005	$\mathbf{0.772 \pm 0.004}$	0.770 ± 0.005
Statics 2011	0.770 ± 0.003	0.822 ± 0.003	0.825 ± 0.002	0.832 ± 0.003	$\mathbf{0.834 \pm 0.002}$
Tutor	0.856 ± 0.003	0.875 ± 0.002	0.882 ± 0.001	0.890 ± 0.0008	$\mathbf{0.895 \pm 0.001}$

The results from table 5 show that qDKT variations outperform DKT on each of the four data sets. For three of the given datasets, qDKT with fastText embedding and regularisation achieves the best score, and for one of the qDKT embedded with fastText achieved the highest AUC. From the given experiments, it can be concluded that various implementations of question-level DKT with a word embedding and regulariser solves some of the issues that DKT has with nested questions under skill. Nevertheless, the proposed qDKT model is relatively new and further experiments are needed to firmly establish its positive effect.

PRACTICAL APPLICATIONS OF DKT

The high efficiency of DKT can be used in many practical ways to improve online educational platforms. Since DKT can produce real-time analysis of what a student knows and does not know, this can allow online educational platforms to dynamically adapt their content and instructions to optimize the depth and efficiency of learning. Traditional self-reporting methods such as questionnaires or interviews cannot capture the temporality of learning processes. Data produced by DKT can be used to detect learning tactics best suited to the student using the online educational platform. In this way, educational websites rely on data-driven methods that can address the needs of individual learners and produce content of the highest quality for the specific user. The task of designing a concrete study plan for a course is no longer only the responsibility of the lecturers, it is rather semi-automated. Online educational systems can use provided learning units to independently produce different study plans based on the student data obtained by DKT.

DKT can also provide an early warning system that predicts subsequent course performance. Prediction data from the model can be utilised to allow online educational platforms to automatically estimate student's confusion. Students experience confusion when they are confronted with an anomaly, contradiction, or an impasse, and are uncertain about how to proceed (Yang, Rose and Kraut, 2016). Struggling with confusion as a cognitive activity may enable learners to acquire a deeper understanding of complex topics. Therefore, DKT is very effective at capturing such cognitive process by looking at past results from students' choices and performance. Confusion is a big problem in online educational platforms because it is often directly associated with lower student achievement. Confusion might transition into frustration, boredom, and ultimate disengagement from the learning process (Larson and Richards, 1991). The more confusion students express and the more they are exposed to other students' confusion, the sooner they drop out of the course (Yang, Rose and Kraut, 2016). By using DKT, data on student confusion can be obtained in real-time, allowing for just-in-time support. Such support allows students to overcome any struggles they might have when not understanding parts of their courses, promoting higher retention. The DKT data can be integrated to provide educators with exact information on the level of student confusion across different courses, giving them a very detailed overview and allowing them to focus on specific parts of their courses where the student confusion is higher. Prediction analysis from DKT enables educators to monitor student progress and to identify at-risk students in advance in order to support early intervention.

Another benefit of DKT is that it can help instructors and educational designers better understand the impact of different learning activities and resources on learners' engagement, and thereby alter curriculum design in the most effective manner. Typically, the task of discovering latent structure or concepts in data is only performed by human experts, taking a significant amount of time and effort. Instead of relying on this strenuous human-centric approach to find important correlation in students' data from educational platforms, DKT prediction data can be used to gain deeper understanding into the most optimal way to format course content.

Overall, the biggest potential impact of the DKT model is in choosing the best sequence of learning items to present to a student. Since the core purpose of online educational platforms is to allow students to learn content presented by a course in an efficient and engaging way, integrating a novel model like DKT with the huge amount of data produced by students using the platforms can have a significant impact on discovering best learning practices. Different performance groups among students generally classify into one of the following four learning tactics: search oriented, content and assessment oriented, content oriented, assessment oriented (Fan et al., 2021). Learning design and learning tactics have clear and consistent pedagogical effect on students' performance depending on their most suitable learning tactic, so having the opportunity to automatically and effectively adjust course structure and presentation with DKT's best sequence of learning items is a vital innovation in the field of online educational platforms.

DISCUSSION

Deep Knowledge Tracing is an exciting and innovative field of deep learning with diverse applications in the field of online educational platforms. It was first developed in 2015. Compared to previous models in the field of Knowledge Tracing such as Bayesian knowledge tracing and Performance factor analysis, DKT relies on state-of-the-art Machine Learning approaches such as Recurrent Neural Networks to effectively predict student knowledge and skills based on past data. This innovative approach can make the model very successful because it allows flexibility of input features and an ability to identify latent categories in items without explicit identification. DKT can infer the effect of exercise order on learning, it can detect degree of skill overlap) and it can use a student's average accuracy up to a certain trial to predict the next trial. Some experiments show that DKT demonstrated an impressive performance advantage over BKT and PFA.

Nevertheless, DKT can also encounter difficulties among the data obtained from educational platforms, possibly hindering the effectiveness of the model. Such data is often scarce, of varying quality, and it is also difficult to compare knowledge data from different institutions. Another potential limitation is with topics that share same skills, as they can have different difficulties, and nested questions. Some of these limitations can be resolved by introducing variation of the original DKT model such as Federated DKT, question centric DKT or DKT with Pre-training Embeddings via Bipartite Graph. However, all these variations are quite new in the field of Knowledge Tracing and their long-term effect has not yet been tested extensively.

One completely novel application of DKT can be integrating the model into coursework assessment structure. Since the effectiveness of DKT can very accurately obtain the knowledge level among all students in a course, this could be a helpful tool for educators preparing coursework. Most educators so far have mostly focused only on the content of a course to prepare assessment exercises, without using any potential data on the students' knowledge level in the process. This preparation strategy can often lead to ineffective coursework assessment as exercises are either too difficult and daunting for most students, or below the level of the majority of students, leading to unrealistically high grades. By having live feedback data on the exact average knowledge level of their students, educators can use this data to prepare assignments and exams that are challenging, but achievable by the top students in the class.

Despite some potential flaws, DKT can have a substantial impact on improving online educational platforms. The model has a lot of potential, as its cutting-edge technology is flexible in processing varying datasets of significant capacity, and it also has the flexibility to be adjusted when needed. The possible effectiveness of DKT brings a lot of practical features to online educational platforms - dynamic content improvement, early warning system to reduce student dropout rate and personalized learning style based on student strengths. Such features are beneficial for both students and educators. Although further research is needed to reinforce the usage of DKT, results from this new and innovative model are very optimistic and reveal that its integration into online educational platforms can make the future of learning brighter for students and educators.

REFERENCES

Anni Sapountzi, Sandjai Bhulai, Ilja Cornelisz, Chris Van Klaveren (2019). Dynamic Knowledge Tracing Models for Large-Scale Adaptive Learning Environments. *International Journal on Advances in Intelligent Systems*, vol 12, pp. 121-129

Bravo-Agapito, J., Romero, S. J., & Pamplona, S. (2021). Early prediction of undergraduate Student's academic performance in completely online learning: A five-year study. *Computers in Human Behavior*, vol 115

Christian Fischer, Zachary a. Pardos, Ryan Shaun Baker, Joseph Jay Williams, Padhraic Yu, Stefan Slater, Rachel Warschauer. (2020). Mining Big Data in Education: Affordances and Challenges. *Review of Research in Education*, vol 44, pp. 130-160

Chun-Kit Yeung, Dit-Yan Yeung. (2018). Addressing Two Problems in Deep Knowledge Tracing via Prediction-Consistent Regularization. *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*, pp. 1-10

Olah, C. (2015). Colah's Blog, 2015. Understanding LSTM networks. [online] Available at https://colah.github.io/posts/2015-08-Understanding-LSTMs/ [Accessed 30 May 2022]

Dr. Wahab Ali. (2020). Online and Remote Learning in Higher Education Institutes: A Necessity in light of COVID-19 Pandemic. *Canadian Center of Science and Education*, vol 10, pp. 16-25

Delgado, F. (2021). Teaching physics for computer science students in higher education during the COVID-19 pandemic: A fully internet-supported course. *The Future Internet*, vol 13

Dumford, A. D., & Miller, A. L. (2018). Online learning in higher education: Exploring advantages and disadvantages for engagement. *Journal of Computing in Higher Education*, vol 30, pp. 252-465

Yang, D., Kraut, R. and Rose, C. (2016) Exploring the Effect of Student Confusion in Massive Open Online Courses, *Journal of Educational Data Mining*, vol 8, pp. 52–83

Gauhar Afshan, Aliya Ahmed, (2020). Distance learning is here to stay: Shall we reorganize ourselves for the post-covid-19 world?. *Anaesthesia, Pain & Intensive Care*, vol 8, pp. 52-83

H. Ishwaran and L. F. James. (2003) Generalized weighted chinese restaurant processes for species sampling mixture models. *Statistica Sinica*, vol 18, pp. 1211–1235

Jesús de la Fuente, Mónica Pachón-Basallo, Flavia H. Santos, Francisco J. Peralta-Sánchez, María Carmen González-Torres, Raquel Artuch-Garde, Paola V. Paoloni and Martha L. Gaetha. (2021). How Has the COVID-19 Crisis Affected the Academic Stress of University Students? The Role of Teachers and Students. *Frontiers in Psychology*, vol. 12

Jinze Wu, Zhenya Huang, Qi Liu, Defu Lian, Hao Wang, Enhong Chen, Haiping Ma, Shijin Wang. (2021). Federated Deep Knowledge Tracing. *WSDM 21*

Kritphong Mongkhonvanit, Klint Kanopka, and David Lang. (2019). Deep Knowledge Tracing and Engagement with MOOCs. *In Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, pp. 340-342

Larson, R. W. AND Richards. (1991). Boredom in the middle school years: Blaming schools versus blaming students. *American Journal of Education*, vol 99, pp. 418-443

Mohammad Khajah, Robert V. Lindsey, and Michael C. Mozer. (2016). How deep is knowledge tracing. *In Proceedings of the 9th International Conference on Educational Data Mining*, pp. 94-102

Pavlik, P., Cen, H., & Koedinger, K. R. (2009). Performance factors analysis: A new alternative to knowledge tracing. *In Proceedings of the 2009 Conference on Artificial Intelligence in Education*, pp. 531–538

S. E. Embretson and S. P. Reise. (2013). Item response theory. *Open Journal of Statistics*, vol 8

Shashank Sonkar, Andrew Lan, Andrew Waters, Phillip Grimaldi and Richard Baraniuk (2020). qDKT: Question-centric Deep Knowledge Tracing. *In Proceedings of The 13th International Conference on Educational Data Mining*, pp. 677 – 681

Pu, S., Yudelson, M., Ou, L, and Huang, Y. (2020). Deep Knowledge Tracing with Transformers. *Artificial Intelligence in Education*, pp.252-256

Yokoyama, S. (2019). Academic self-efficacy and academic performance in online learning: A mini review. *Frontiers in Psychology*, vol. 9

Yunfei Liu, Yang Yang, Xianyu Chen, Jian Shen, Haifeng Zhang and Yong Yu. (2020). Improving knowledge tracing via pre-training question embeddings. *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, pp 1556-1562

Y. Goldberg and O. Levy. (2014). word2vec explained: deriving Mikolov et al.'s negative-sampling word-embedding method. arXiv preprint

Yizhou Fan, Wannisa Matcha, Nora'ayu Ahmad Uzir, Qiong Wang, Dragan Gasevic. (2020). Learning analytics to reveal links between learning design and self-regulated learning. *International Journal of Artificial Intelligence in Education*, vol 31, pp. 980-1021

Lipton, Z. C., Berkowitz, J., and Elkan, C. (2015). A critical review of recurrent neural networks for sequence learning. arXiv preprint