



ONTOLOGICAL AUGMENTATION AND ANALYTICAL PARADIGMS FOR ELEVATING SECURITY IN HEALTHCARE WEB APPLICATIONS

NAWAF ALHARBE*

Abstract. The 4th and 5th industrial revolutions provided huge access to a high volume of information in healthcare sectors. However, the vast majority of web applications are under attack constantly. Due to the huge volume of attack vectors in recent years, technical breaches to the web applications of healthcare sectors are becoming a common issue. Hence, it is essential to develop an effective framework that would help experts and healthcare practitioners in web application security management. In this research, we proposed an ontology-based technique for developing secure online web applications in healthcare sectors. This study presents a conceptual framework with 5 stages namely: idea understanding, requirement identification, design & code, threat classification, and facilitate. The proposed methodology involves several various advantage features such as providing a unified path for future professional applications. Also, the proposed solution provides a clear pathway for implementing the easy-to-use secure web application development in short term in the healthcare sector. Finally, the study used a multi-criteria decision-making methodology to undertake a performance simulation assessment. Out of the evaluated process, Fuzzy-AHP has shown better performance for scheduling each step of the final developed security system better. In the future, the proposed method will be evaluated using more databases in the healthcare sector to improve the applications of the proposed method.

Key words: Security, decision support system, healthcare web application, machine learning application

AMS subject classifications. 68T01, 68U35

1. Introduction. Virtual platforms and the Internet of Things (IoT) have been two-man pillars of the fourth and fifth revolutions [27]. Industry 4 provided easy access to the vast volume of information. New generations of wireless networks such as the fifth generation (5G) and future generation sixth generation (6G) will provide faster and more reliable to process this information [4]. Nowadays there are sensors and intelligent devices that monitor human daily activities and store information. In this environment, collected data is processed through cloud computing devices and collected data by IoT devices are uploaded to cloud servers [32]. Healthcare is one of the industries that benefited from various sources of information and the availability of datasets to provide a portfolio for each patient. By increasing the applications of IoT, the chance for adversaries to steal, and alter patient information increases as well. The healthcare sector is compacted with vital information and vast amounts of data in form of biomedical signals and patients' treatment histories. Due to the intricacies of such information, maintaining and securing the entire healthcare data is essential for any healthcare web application [2].

The main focus of these attacks is the exploitation of healthcare web apps. The prevalence of these attacks has increased from 2005 to 2022. In total, 256.65 million breach attempts have been recorded from 2005 to 2021 [16]. the total number of attacks registered and implemented with regulating authorities is 3868 from all over the world, with 627 breaches in the preceding year [23]. The number of breaches in the healthcare sector depicts the vulnerability of this industry against such an attack. IBM reported due to the swarm of attacks against the healthcare sector and the correspondent breaches caused 6.45 6.45 million dollars in loss for this industry [16]. Due to the prevalence of attacks and breaches in the healthcare section, securing the developed web application is the utmost pressing need.

Another major consequence of these breaches to the database is corresponding harm to the company's reputation. Customers don't want their online applications to be interrupted or lose data while using them [26, 25, 33]. Most of the time, the developer overlooks the possible faults and vulnerabilities since the main priority of the developer is providing an application with optimal usability over the estimated deadline [22, 31].

*Applied College, Taibah University, Medina 42353, Saudi Arabia (nrharbe@taibahu.edu.sa).

Furthermore, many firms are unable to determine the vulnerability of the proposed web application before the breaches occur.

To address the presented issues in this article, the authors proposed a solution for online web applications in the healthcare sector using a combination of AI semantics and formal ontologies. When it comes to web application security, we must consider the entire development functionality of the application. Security in web applications can only be addressed by an effective and improved security approach during the development process [23, 3]. A well-equipped team and laboratory are required to produce a secure development toward structured security of web application. From an ontology standpoint, the projected study aims to give a methodical, step-by-step approach for designing secure healthcare online applications. Ontology is a new effective concept or idea that has a lot of promise for displaying solid and successful results. The projected framework identifies the security factors that must be addressed from the beginning of the web application development process and then handled in a systematic manner [13].

The projected framework is a conceptual ontology-based approach aimed at providing efficient and secure online application development. This project creates a pathway for repositories to provide the best professional practices and processes to enable secure development in healthcare web applications. The advancement of security necessitates its application at the earliest stages of development. It is necessary to discuss and provide an artifact-based method that produces an effective result to make this practicable. This study is organized into six main sections. In Sect. 2, the history of the ontology-based method for intrusion detection and recently proposed solutions are discussed. Section 3 describes the importance of the problem and corresponding statistical characteristics related to it. In Sect. 4, the proposed model and strategy are presented. Section 5 evaluated the proposed model and verified the achieved results. Finally, Sect. 6 summarizes the research contributions, limitations, and future of the work.

2. Related Works. The fourth industrial revolutions publicized the prevalence of gadgets and electronic devices in various industries. Healthcare is one of these industries with a huge volume of electronic records and contextual information [30]. However, the quality and quantity of the database have increased recently, and the number of attacks and variety of attack types has increased respectively [3]. Securing vast amounts of information and communications through the virtual platform is very important. Artificial Intelligence (AI) applications have shown huge potential for intrusion detection and improving the security of the web base applications [19]. Due to the vast majority of information, Ontology-based security frameworks and their tools can benefit from semantic analysis and its tools [24]. However, developing a proper ontology-based strategy for the healthcare section is difficult. In this section, we reviewed related research to securing healthcare web applications. The following studies have reviewed the effective solution for secure healthcare web applications using semantic ontology-based methods. Due to the recent prevalence of smart devices, the environment around the patients or even inside the patients' bodies is filled with these instruments.

The process of securing recording and transferring information for patients is vital. Any misleading information in this database might cause fatality for patients [5]. Henaien et al. [5]. Presented a method based on 4 stages of ontology to manage user profile information, medicine prescription data, sensor information, and identification process. To secure the collected information, the proposed method follows the security process based on the patient's personalization information. If the patient doesn't suffer from memory, optical, or hearing impairments then they can access all of the recorded information via passwords. Otherwise, caregivers, doctors, or other third parties must help the patients to access this information.

Chen et al. [9]. Proposed an ontology-based framework in 4 layers. The presented method involves a service discovery and selection layer, a business layer, a data access layer, and a technological layer. Each of the presented layers operates to satisfy certain trends. The data access layer manages the database synchronization, business layer manages the service operation system. The proposed security layer uses the Artificial Intelligence (AI) semantic mediator to choose the user desired web service. The proposed method tried to create a secure portfolio using an integrated medical and healthcare dataset.

Sharma et al. [29]. Presented an ontology based on AI to monitor Biomedical Signals such as ECG, PPG, temperature, heart rate, and HR Variability. The authors collected the mentioned information using available gadgets such as apple watch and bio best. The authors used artificial intelligence methods such as K Nearest Neighbor (KNN) [34] to detect the COVID-19 prevalence in each cohort. The authors concluded that

the proposed ontology can detect the prevalence of COVID-19 among each cohort with 96.33% accuracy. The process of sharing information via each client is edge and server base sharing with the lowest consumption of energy. The authors concluded that the proposed method can be used for safe monitoring among COVID-19 patients via a 2-layers security transferring process.

Hady et al. [14]. Presented a combination of network and biometric metrics as features for intrusion detection. The authors used Support Vector Machine (SVM) [7], KNN, and artificial neural network [1] to detect the intrusion. The authors gathered biomedical information such as blood oxygen, temperature, and heat bit rate while recording network information such as computer identification data. The proposed program detects intrusion attacks such as spoofing attacks and data alteration. The authors created a database using 16 thousand collected records from various patients. The authors reported 92.44% accuracy using SVM for intrusion detection.

Mozzaquatro et al. [20]. Presented an ontology-based solution using the combination of semantic ontological and formal solutions. The authors presented solution monitories the reliability, operability, operability, transferability, and adaptability characteristics of each web application. The authors concluded that the proposed strategy reduces risk and ensures long-term software serviceability and security. The presented solution by the authors in this research provides the reliable automated automatic establishment of security metrics using explicit and reasoning information about situations of interest and combined knowledge from multiple security experts.

Bataityte et al. [6]. Presented a solution based on three layers of formalization to automate the formal solutions to solve the cyber security problem such as logical vulnerability, and risk assessment. The presented solution enables analysing some cyber security logical problems such as vulnerability analysis and risk assessment. The authors presented a logical theory based on situations and actions in descriptive logical theory. The authors deployed the presented solution into the ontology web language, and semantic web rule language [21]. The authors concluded that the proposed solution can provide simple analytic results related to logical vulnerability, risk assessment, and policy validation.

Shahzad et al. [28]. Presented an ontology base solution to combine information of integrated health care intelligent devices without any security problem. The proposed solution is composed of three layers of initial health system analysis, proposing ontological representation of the smart system and finally integrating every system. The proposed method is evaluated on arrhythmia detection, prostate cancer care process, and leukaemia detection. In each use case, the presented solution gathered biomedical information from each patient then detect the related disease, and finally reported the problem to caregivers or correspondent doctors. Finally, the proposed solution recorded the information about each event for each patient.

The reviewed articles attempted to fully comprehend their work before presenting a suitable framework with unique utilities and working techniques. However, the presented solution didn't consider the effect of simplicity and easy development for healthcare sectors. Adopting ontology during the development of secure applications of healthcare is critical because there is currently no specialized work accessible that discusses web-based application security from the perspective of health informatics, or slightly another web application perspective.

Another important issue is the process of validating the presented solution by ontology. Healthcare sector web applications are integrated and the concept of integration must be a crucial point to develop and test the proposed ontology solution. However, the majority of the proposed solution evaluated the ontology only in the specific domain of the presented solution, not as an integrated web application solution.

To address these issues the proposed solution by this article investigates the ontology base solution to secure the web applications in the healthcare sectors in a simple process. The proposed solution is evaluated to other solution to prove the superiority of the presented integrated ontology over similar research.

3. Problem Statement. As mentioned previously, the number of developed applications in the healthcare section has increased in the last decades. The result of the ongoing increase in the quantity and quality of the available data is the rise in the number of attacks. Table 3.1 represents the number of recorded attacks and their relation to the total volume of datasets.

The represented information in Table 3.1 is based on summarized reports from HIPPA [3]. Based on the reported information from Table 3.1, the number of reported breaches has increased 5 times over the past

Table 3.1: Number of recorded attacks based on available datasets.

Year	Amount of Data (in Millions)	Number of recorded attacks
2010	5.5300	199
2011	13.150	200
2012	2.8000	217
2013	6.9500	278
2014	17.450	314
2015	113.27	269
2016	16.400	327
2017	5.1000	359
2018	33.200	365
2019	41.200	505
2020	52.350	835
2021	65.540	987

Table 3.2: Attacks Source statistics

Year	Technical Issues and Causes	Human Error and Disclosure	Theft	Miss Handling
2010	8	8	148	10
2011	17	27	136	7
2012	16	25	138	8
2013	25	64	150	13
2014	35	76	143	12
2015	57	101	105	6
2016	113	129	78	7
2017	147	128	73	11
2018	158	143	55	9
2019	274	142	51	7
Total	850	843	1077	90

decades. Because health is inextricably linked to human life, the number of breach records is quite noteworthy and high. A data leak of this nature can put patients' lives in jeopardy. Furthermore, it is to mend its flaws after studying the attack statistics in healthcare. It is vital to initially discover the gaps in healthcare breach occurrences to identify the key difficulties. Detailed information about each category of attack has shown in Table 3.2.

Table 3.2 displays breach' counts carried out by each source, which provides a clear point-by-point explanation of healthcare flaws and aids professionals in preparing preventive actions. Based on the presented categories of attacks in Table 3.2, the maximum typical and frequent exploit generators in the healthcare sector are technical issues and human errors. Although, when comparing Table 3.1 and Table 3.2, it is clear that on-line application vulnerabilities generate the most troubles and exploits in healthcare. Technical flaws produce and generate troubles in the modern era, and attackers take advantage of these flaws by injecting healthcare applications on a wide scale regularly. The data and statistics presentation in Table 3.1 and Table 3.2 depict the critical issue of online medical system management, and also the necessity for functionality-centric security solutions. This research proposed an ontology-based framework to provide specialists with greater affectivity and a conceptual grasp of how to design healthcare online applications.

4. Methodology. Ontology is widely accepted and applied in various domains such as communication development as well as actual interaction between humans and machines [15]. The ontology adaption is a novel concept from the perspective of healthcare web application development. The ontology-based idea can be

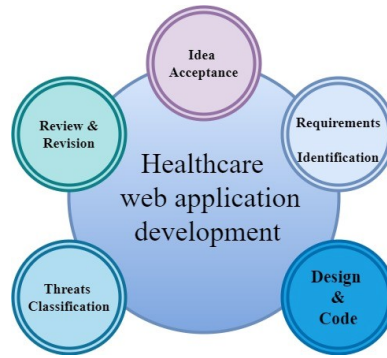


Fig. 4.1: Projected framework.

used to develop intergraded web applications for an operative as well as efficient security purposes. Ontology helps the professional to translate real-world problems into machine settings and solve these problems using the machine. However, utilizing any ontology-based solution for securing healthcare virtual platforms leads to more complications in the process. Another important problem issued in this research is the vast quantity of data obtainable on many online as well as offline sites which developers accept and use without any background check to make their work easier [17].

In this article, we proposed an ontology-based framework to effectively secure the development of healthcare web applications. The proposed framework merges capable ontology security management with the traditional and commonly used development stages. The proposed methodology is created using the adopted ideology and can be easily developed in real-world web applications. Any new approach must first develop a notion before it can be fully addressed. In this work, we presented a standard development paradigm with ontology. The proposed framework has shown in Fig. 4.1.

Figure 4.1 depicts a five-step integrated systematic framework for securing applications at the early stages of the development process that combines traditional development processes with security management features. The proposed paradigm combines concept initiation with ontology to help with the development phase of the web application.

The proposed framework uses 5 phases to to secure the web application based on the developer needs. These five phases are: Requirement Identification; Idea Understanding; Classification of Potential Possible Threats; Design & Code, and facilitate. To make the process easier, no additional specific stages are there in the process that would demand extra efforts or time from the developer or the organizations. We attempted to keep things basic and straightforward because complication breeds both confusion and a lack of enthusiasm for the development process on the developer's part. framework's step-by-step working method, which is explained topic of this study is mentioned as follows:

- **Idea Understanding:** This stage gets people thinking about progress. In this segment, the web application's programmer, developer, as well as owner, discuss the functionality and ideas for how the web application will look after development, as well as what kinds of features the owner wants.
- **Requirement Identification:** This stage determines which processes, utilities, and attributes need to be handled during the program's design and development. The need identification phase integrates these demands and creates a good new framework for the application's systematic development.
- **Design & Code:** Designers design the code that correlates the stated necessities and create an effective application based on them in this step.
- **Threat Classification:** This is the phase in the process that protects the system from exploitation. This is a process in which three viewpoints are analysed and then the existing development scenario is compared. If the scenario is compatible with the demand, it's acceptable; otherwise, the loop begins over with the requirement identification procedure.
- **Facilitate:** The projected framework permits designers to assist the developed web application in the

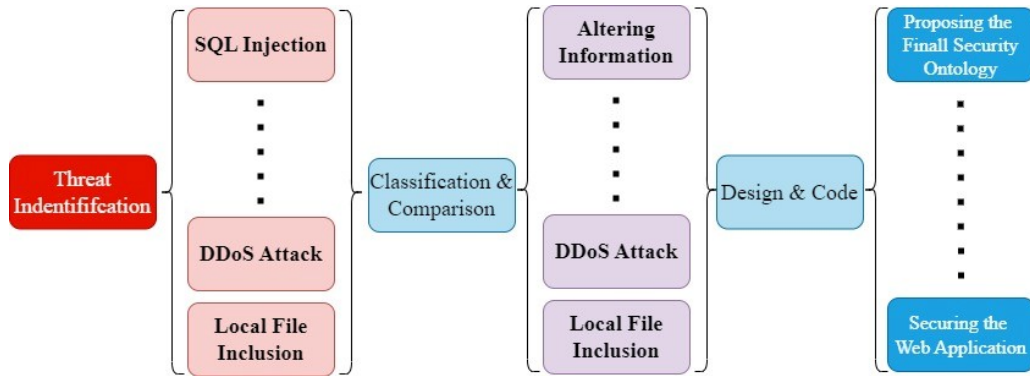


Fig. 4.2: Projected framework taxonomy.

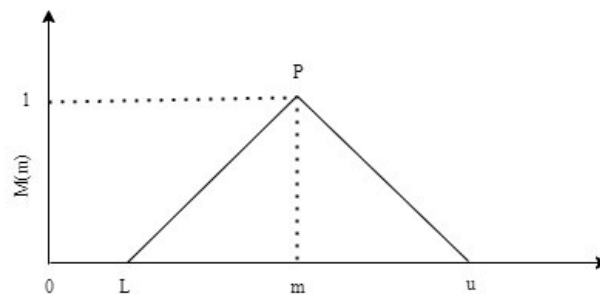


Fig. 5.1: Triangular fuzzy number (P).

healthcare industry after classifying and comparing the numerous threats and other development issues. The anticipated framework allows the designers to simplify the generated web application in the healthcare market once the full cycle becomes systematic and uninterrupted.

Various steps of identification, classifying the threats, and final design of the proposed taxonomy has shown in Fig. 4.2. Overall, after examining the framework, concept, and working flow, it is evident that the presented framework is the first stage in using an ontology-based method to assist healthcare web application development. The projected framework depicts the authors' conceptual framework and will require further mathematical testing and confirmation in the future

4.1. Experimental setting. In this research, the proposed method is compared with benchmark frameworks to determine its efficacy of it. We used six distinct frameworks namely OWASP Ontology Framework (FW1) [12], Moreira Framework (FW2) [10], Mozzaquatro Framework (FW3) [24], Maglogiannis Framework (FW4) [8], and Analytical Hierarchy Process mixed with fuzzy set theory (Fuzzy-AHP) (FW5) [18]. Fuzzy-AHP is a method that produces accurate, widely accepted, and validated outcomes. The projected framework in this research is Fuzzy-AHP.

5. Experimental Result. To compare the result, the triangular fuzzy number is used and the result has shown in Table 5.1.

We used a cut strategy to aggregate the fuzzy set into the values after detecting the pair-wise fuzzy numbers [11]. The triangular membership function has shown in Fig. 5.1.

As shown in Fig. 5.1, the Parameter L represents the minimal value, parameter m represents the most likely value, and parameter u represents the maximum value. The membership equation are shown in the

Table 5.1: Attacks Source statistics.

Model Name	FW1	FW2	FW3	FW4	FW5
OWASP Ontology Framework (FW1)	1.00000, 1.000000, 1.000000	0.69000, 0.886000, 1.100000	0.226000, 0.276000, 0.357000	1.000000, 1.516000, 1.933000	0.490000, 0.637000, 1.000000
Moreira Framework (FW2)	*	1.00000, 1.000000, 1.000000	0.695000, 0.950000, 1.346000	0.268000, 0.352000, 0.518000	0.166000, 0.197000, 0.253000
Mozzaquatro Framework (FW3)	*	*	1.000000, 1.000000, 1.000000	1.000000, 1.320000, 1.5520008	0.301000, 0.435000, 0.803000
Maglogiannis Framework (FW4)	*	*	*	1.000000, 1.000000, 1.000000	0.222000, 0.287000, 0.415000
Projected Framework (FW5)	*	*	*	*	1.000000, 1.000000, 1.000000

Table 5.2: Attacks Source statistics.

Model Name	FW1	FW2	FW3	FW4	FW5
OWASP Ontology Framework (FW1)	1.000000	1.260900	1.610200	1.002100	0.902100
Moreira Framework (FW2)	0.788000	1.000000	1.269000	0.660500	0.505000
Mozzaquatro Framework (FW3)	0.602000	0.780800	1.000000	0.650400	0.690000
Maglogiannis Framework (FW4)	0.907900	1.500400	1.530000	1.000000	0.606500
Projected Framework (FW5)	1.080700	1.810700	1.440900	1.500500	1.000000

following equations.

$$P(x|m) = \{0; x < L\} \quad (5.1)$$

$$P(x|m) = \left\{ \frac{x-L}{m-L}; L \leq x < m \right\} \quad (5.2)$$

$$P(x|m) = \left\{ \frac{u-x}{u-m}; m \leq x < u \right\} \quad (5.3)$$

$$P(x|m) = \{0; x > u\} \quad (5.4)$$

Table 5.2 shows the cut approach's value as well as the aggregated value of the triangular fuzzy numbers.

Based on what is described in Table 5.1, and aggregated information in Table 5.2, the projected frameworks demonstrates better option compare to other methods.

As shown in Fig. 5.2, the priority for the evaluated framework is $FW5 > FW4 > FW1 > FW3 > FW2$. The findings of the performance simulation mentioned in this part show that FW5 has the highest effectiveness and priority, whereas FW2 has the lowest. In general, the presented framework is one of the most appropriate

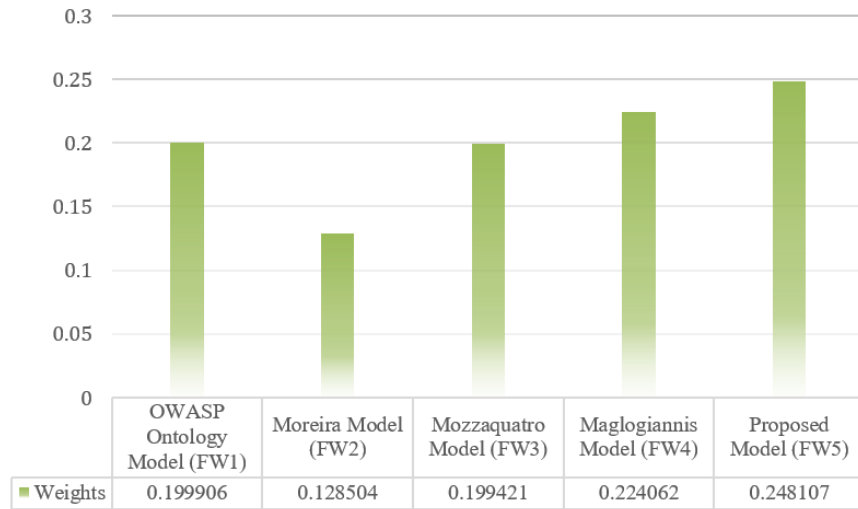


Fig. 5.2: Graphical illustration of priority.

frameworks and ideologies that academics and future practitioners in their subject might follow. The proposed framework in this section compared the performance of each section and provided a systematic workflow to effectively secure the web application in the healthcare sector. This section depicts an optimal pathway to simplify the difficulty of proposing a systematic and statistically assessed efficacy order. The projected framework has the highest effect ratio in all of the selected frameworks, with a weight of 0.248107 and a CR value of 0.001547.

6. Conclusion. Due to the development of more virtual applications and the increased volume of the dataset, the number of attacks on healthcare web applications has increased. Any breach the healthcare web applications due to the vast number of attacks lead to losing vital information and harm to the company's brands. To address these issues, we proposed an easy-to-use pre-security procedure to ensure the healthcare security section. The presented methodology is an ontology-based solution to check the possible threats and develop a proper design for the proposed framework before releasing it to the public. The main focus of this projected framework is to provide an appropriate path for future researchers. The contributions of the proposed framework are: 1. Proposing the requirement of a secure web application platform in the healthcare section. 2. Providing comparison and classification path for the proposed security program. 3. Providing a novel pathway for future ontology-based security solutions in the healthcare sector. In the future, the proposed solution will be evaluated using more experimental analysis to provide a general security application in the healthcare domain.

REFERENCES

- [1] O. I. ABIODUN, A. JANTAN, A. E. OMOLARA, K. V. DADA, A. M. UMAR, O. U. LINUS, H. ARSHAD, A. A. KAZAURE, U. GANA, AND M. U. KIRU, *Comprehensive review of artificial neural network applications to pattern recognition*, IEEE Access, 7 (2019), pp. 158820–158846.
- [2] I. ABU-ELEZZ, A. HASSAN, A. NAZEEMUDEEN, M. HOUSEH, AND A. ABD-ALRAZAQ, *The benefits and threats of blockchain technology in healthcare: A scoping review*, International Journal of Medical Informatics, 142 (2020), p. 104246.
- [3] M. ALENEZI, *An ontological framework for healthcare web applications security*, International Journal of Advanced Computer Science and Applications, 12 (2021).
- [4] R. S. ALONSO, J. PRIETO, F. DE LA PRIETA, S. RODRÍGUEZ-GONZÁLEZ, AND J. M. CORCHADO, *A review on deep reinforcement learning for the management of sdn and nfv in edge-iot*, in 2021 IEEE Globecom Workshops (GC Wkshps), IEEE, 2021, pp. 1–6.
- [5] O. T. AROGUNDADE, A. ABAYOMI-ALLI, AND S. MISRA, *An ontology-based security risk management model for information systems*, Arabian Journal for Science and Engineering, 45 (2020), pp. 6183–6198.

- [6] K. BATAITYTE, V. VASSILEV, AND O. J. GILL, *Ontological foundations of modelling security policies for logical analytics*, in Artificial Intelligence Applications and Innovations: 16th IFIP WG 12.5 International Conference, AIAI 2020, Neos Marmaras, Greece, June 5–7, 2020, Proceedings, Part I 16, Springer, 2020, pp. 368–380.
- [7] J. CERVANTES, F. GARCIA-LAMONT, L. RODRÍGUEZ-MAZAHUA, AND A. LOPEZ, *A comprehensive survey on support vector machine classification: Applications, challenges and trends*, Neurocomputing, 408 (2020), pp. 189–215.
- [8] V. CHARPENTIER, N. SLAMNIK-KRIJESTORAC, AND J. MARQUEZ-BARJA, *Latency-aware c-its application for improving the road safety with cam messages on the smart highway testbed*, in IEEE INFOCOM 2022-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), IEEE, 2022, pp. 1–6.
- [9] S.-W. CHEN, Y.-T. TSENG, AND T.-Y. LAI, *The design of an ontology-based service-oriented architecture framework for traditional chinese medicine healthcare*, in 2012 IEEE 14th International Conference on e-Health Networking, Applications and Services (Healthcom), IEEE, 2012, pp. 353–356.
- [10] J. R. DAVIS, *Healthcare Entities and Data Breach Threat Indicators and Deterrence: A Quantitative Study*, PhD thesis, Northcentral University, 2022.
- [11] E. DOS SANTOS MOREIRA, L. ANDRÉIA FONDAZZI MARTIMIANO, A. JOSÉ DOS SANTOS BRANDÃO, AND M. CÉSAR BERNARDES, *Ontologies for information security management and governance*, Information Management & Computer Security, 16 (2008), pp. 150–165.
- [12] K. N. DURAI, R. SUBHA, AND A. HALDORAI, *A novel method to detect and prevent sqlia using ontology to cloud web security*, Wireless Personal Communications, 117 (2021), pp. 2995–3014.
- [13] M. GHEISARI, H. E. NAJAFABADI, J. A. ALZUBI, J. GAO, G. WANG, A. A. ABBASI, AND A. CASTIGLIONE, *Obpp: An ontology-based framework for privacy-preserving in iot-based smart city*, Future Generation Computer Systems, 123 (2021), pp. 1–13.
- [14] A. A. HADY, A. GHUBAISH, T. SALMAN, D. UNAL, AND R. JAIN, *Intrusion detection system for healthcare systems using medical and network data: A comparison study*, IEEE Access, 8 (2020), pp. 106576–106584.
- [15] I. HARROW, R. BALAKRISHNAN, E. JIMENEZ-RUIZ, S. JUPP, J. LOMAX, J. REED, M. ROMACKER, C. SENGGER, A. SPLENDIANI, J. WILSON, ET AL., *Ontology mapping for semantically enabled applications*, Drug discovery today, 24 (2019), pp. 2068–2075.
- [16] T. HOU AND V. WANG, *Industrial espionage—a systematic literature review (slr)*, computers & security, 98 (2020), p. 102019.
- [17] C. ISLAM, M. A. BABAR, AND S. NEPAL, *An ontology-driven approach to automating the process of integrating security software systems*, in 2019 IEEE/ACM International Conference on Software and System Processes (ICSSP), IEEE, 2019, pp. 54–63.
- [18] Y. LIU, C. M. ECKERT, AND C. EARL, *A review of fuzzy ahp methods for decision-making with subjective judgements*, Expert Systems with Applications, 161 (2020), p. 113738.
- [19] B. K. MOHANTA, D. JENA, U. SATAPATHY, AND S. PATNAIK, *Survey on iot security: Challenges and solution using machine learning, artificial intelligence and blockchain technology*, Internet of Things, 11 (2020), p. 100227.
- [20] B. A. MOZZAQUATRO, C. AGOSTINHO, D. GONCALVES, J. MARTINS, AND R. JARDIM-GONCALVES, *An ontology-based cybersecurity framework for the internet of things*, Sensors, 18 (2018), p. 3053.
- [21] A. PATEL AND S. JAIN, *Present and future of semantic web technologies: a research statement*, International Journal of Computers and Applications, 43 (2021), pp. 413–422.
- [22] I. RUS, M. LINDVALL, AND S. SINHA, *Knowledge management in software engineering*, IEEE software, 19 (2002), pp. 26–38.
- [23] K. SAHU, F. A. ALZHRANI, R. SRIVASTAVA, AND R. KUMAR, *Hesitant fuzzy sets based symmetrical model of decision-making for estimating the durability of web application*, Symmetry, 12 (2020), p. 1770.
- [24] K. SAHU AND R. SRIVASTAVA, *Soft computing approach for prediction of software reliability*, Neural networks, 17 (2018), p. 19.
- [25] ———, *Needs and importance of reliability prediction: An industrial perspective*, Information Sciences Letters, 9 (2020), pp. 33–37.
- [26] M. SAITO, A. HAZEYAMA, N. YOSHIOKA, T. KOBASHI, H. WASHIZAKI, H. KAIYA, AND T. OHKUBO, *A case-based management system for secure software development using software security knowledge*, Procedia computer science, 60 (2015), pp. 1092–1100.
- [27] N. I. SALEH, M. T. IJAB, AND N. HASHIM, *A review on industrial revolution 4.0 (ir4. 0) readiness among industry players*, in International Conference on Computer, Information Technology and Intelligent Computing (CITIC 2022), Atlantis Press, 2022, pp. 216–231.
- [28] S. K. SHAHZAD, D. AHMED, M. R. NAQVI, M. T. MUSHTAQ, M. W. IQBAL, AND F. MUNIR, *Ontology driven smart health service integration*, Computer Methods and Programs in Biomedicine, 207 (2021), p. 106146.
- [29] N. SHARMA, M. MANGLA, S. N. MOHANTY, D. GUPTA, P. TIWARI, M. SHORFUZZAMAN, AND M. RAWASHDEH, *A smart ontology-based iot framework for remote patient monitoring*, Biomedical Signal Processing and Control, 68 (2021), p. 102717.
- [30] A. TRAUTH-GOIK, *Repudiating the fourth industrial revolution discourse: a new episteme of technological progress*, World Futures, 77 (2021), pp. 55–78.
- [31] O. I. UWAKWEH, *Cybersecurity in the Retail Industry: Third Party Implications*, PhD thesis, University of Cincinnati, 2020.
- [32] G. WANG AND F. XU, *Regional intelligent resource allocation in mobile edge computing based vehicular network*, IEEE Access, 8 (2020), pp. 7173–7182.
- [33] S.-F. WEN AND B. KATT, *An ontology-based context model for managing security knowledge in software development*, in 2018 23rd Conference of Open Innovations Association (FRUCT), IEEE, 2018, pp. 416–424.
- [34] L. XIONG AND Y. YAO, *Study on an adaptive thermal comfort model with k-nearest-neighbors (knn) algorithm*, Building and Environment, 202 (2021), p. 108026.

Edited by: Chiranji Lal Chowdhary

Special issue on: Scalable Machine Learning for Health Care: Innovations and Applications

Received: Feb 22, 2023

Accepted: Dec 8, 2023



TRADITIONAL CULTURAL NETWORK ONLINE EDUCATION INTEGRATING DEEP LEARNING AND KNOWLEDGE TRACKING

HENG ZHAO* AND ZHIYUAN SUN†

Abstract. With the popularization of the computer network and the development of artificial intelligence technology, the traditional education industry has been reformed. In the past two years, online education has developed rapidly. The combination of the Internet and education enables students to study online at any time, no longer relying on the time and place requirements in traditional education. However, with the rapid development of online education, many problems have gradually emerged. In online education, with a large amount of knowledge and question banks, students are faced with a large number of choices. Therefore, positioning and tracking the knowledge level of students and realizing personalized online education have become the main problems facing the moment. Based on this, this study integrates deep learning and knowledge tracking technology to build a traditional cultural network online education model, aiming at accurately positioning students' knowledge levels and recommending personalized question banks. The experimental results show that the average AUC of the model proposed in this study is 0.781, and the average accuracy rate is 0.886, which is significantly better than other online education models. Through the combination of deep learning and knowledge tracking technology, the research successfully provides a new and efficient model for personalized learning in the field of online education, which is of great guiding value for promoting further innovation and development of online education. In addition, the research also provides practical solution strategies for related fields, which have obvious practical significance and popularization value. Online education; Personalized recommendation; Deep learning; Knowledge tracking

Key words: Online education; Personalized recommendation; Deep learning; Knowledge tracking cost

1. Introduction. As the popularization of the Internet and the growth of artificial intelligence technology, online education has gradually become one of the trends of modern education [1]. With the advantages of the Internet and big data, online education has changed the demand for time and location in traditional education [2]. And it can provide users with rich, real-time updated learning materials, reducing the gap between the education level of different development levels in the region [3]. However, with the gradual development of online education, more problems are gradually revealed. These problems are as follows: teachers can not follow up the learning status of students in real time; students in the knowledge base is difficult to choose; the increase of meaningless learning and other problems. The current online education industry practitioners are plagued by these problems [4]. In the traditional teaching mode, teachers are difficult to rely on manual modeling of learning resources for each student to achieve the screening of learning, and they are also unable to accurately perceive the learning changes of learners. As the advancement of artificial intelligence technology, each learner can easily build a unique learning resource model. However, how the huge database of teaching resources can be accurately recommended to students to achieve their personalized development is still a major problem that needs to be solved [5]. To address this problem, this study proposes a traditional culture network online education model that integrates deep learning (DL) and knowledge tracking (KT), aiming to discover learning resources suitable for learners from a large amount of data through DL. And through KT technology, students' learning data are mined and used for online education, aiming to improve the correct rate of predicting students' correct answers to questions, thus improving the effectiveness of online teaching. In addition, the research innovatively combines knowledge mapping to obtain potential connections between items through the description of semantic associations. The research also expands users' interests through various types of associative relationships, which improves the accuracy and diversity of traditional online education recommendation systems. The research aims to combine modern science and technology to provide more effective aids for online teaching and to help learners learn more efficiently. The research also provides a

*School of Tourism and Hotel Management, The Open University of Shaanxi, Xi'an, 710119, China (hengzhaohz@outlook.com)

†School of Marxism, The Open University of Shaanxi, Xi'an, 710119, China (xuegen2021@163.com)

reference for the application of DL and KT in different fields. This research is composed of five main parts. The first part is an overview of the research as a whole; the second part is a summary of related work at home and abroad; the third part is divided into two subsections; the first one describes the model improvement of KT algorithms based on DL, and the second one describes the construction of KT online education models based on DL; the fourth part is an experimental validation of the performance of the KT models proposed in the study; the fifth part is a summary of this study, and it has been an outlook on the future research.

2. Related works. Recently, research on DL has attracted wide attention in various fields. Shen L et al. believed that the short-term forecast of passenger flow had an important significance to the scheduling of subway systems and other aspects. An improved gravity model with DL was proposed, and a short-term forecasting method was upgraded to balance model interpretability and forecasting accuracy. This method had good performance in the experimental results, which was better than other models [6]. This method provided a more reliable solution for improving the accuracy of short-term forecasting of passenger flow in subway systems. In the data-driven decision support system, Garg et al. designed a new data processing framework based on a DL network by using the combination of density-based clustering and variable-length sequence decoding method. The concept of data association was used to track the occurrence and evolution of such events to understand the operating environment. Simulation experiment denoted that this method could perceive the sensitivity of performance and parameters better than others [7]. The method provided a new reliable reference for DL in data processing. Tang et al. found that the linear defect detection of a photovoltaic module was a key link in the health assessment of photovoltaic devices, but the traditional defect diagnosis was actually inefficient manually. Therefore, by distributing computing tasks among edge devices, edge servers and cloud servers, a defect detection algorithm was proposed based on DL. The effectiveness and accuracy of the method were verified by experiments [8]. The method simplified the task of allocating computation between edge and cloud servers for big data processing. Bernardini et al. introduced a DL framework to predict baryon fields. The new model included two network parts, U-Net and Wasserstein, and combined the universe volume and amplified fluids from the real-life environmental feedback project. The dynamical simulation was used to represent a large range of scales. The experimental results indicated that the accuracy was within 10%, which was in good agreement with the cosmological simulation [9]. The research introduced DL to cosmological tasks, providing a viable solution for baryon field prediction. Liu C et al. developed a hierarchical approach based on DL network embeddings to identify patient subtypes from large-scale patient somatic mutation profiles. A network embedding approach encoded genes on the protein interaction set to construct patient vectors. The high classification accuracy indicated that the web-embedded based patient features were reliable in classifying patients [10]. This study provided a powerful DL approach for personalized cancer treatment.

As online education has become a popular research field, the research on data mining technology in online education has also attracted the attention of scholars, and the KT model has become a key technology for simulating the state of students. Song X et al. believed that there were problems in KT research that considered the single relationship and the lack of interpretability, so they proposed a deep KT framework based on a joint graph convolutional network, which modeled multi-dimensional relationships as graphs for relationship fusion. Experiments proved that the method performed better than others. And the interpretability of learning analysis was proved through a case study [11]. The study provided a more effective solution to the problem of considering a single relationship with a lack of interpretability. Pavlik PI et al. introduced a formal learner modeling method - logical knowledge tracing (LKT), which integrated many existing learner modeling methods. The advantage was that the logistic regression model of choice provided the specification of a symbolic system that could specify many existing models in the literature and many new ones. Experimental outcomes indicated that the approach considering multiple learner model characteristics and learning environments was correct [12]. This study integrated existing learner modeling approaches and provided an easier reference for future research. To alleviate the sparsity problem in the summary KT online learning system, Gan W et al. proposed an enhanced learning model with an attention mechanism for graph representation, and used the model to skill embeddings. Experiments from three datasets verified the superiority and interpretability of this model [13]. This research provided a reliable solution to the sparsity that exists in online learning systems. Huang Y et al. found that it was difficult for students to find suitable exercises from a large number of topics provided by many online systems in programming training. This study found a new model for KT, which added additional information

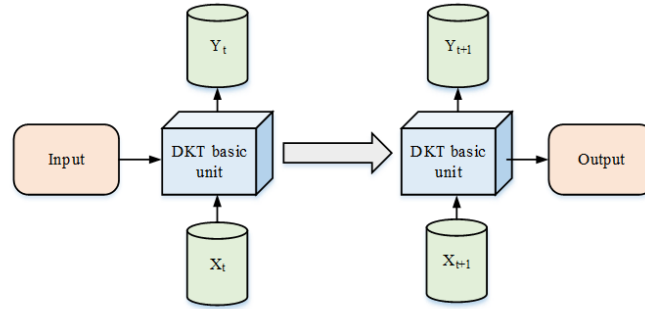


Fig. 3.1: DKT model structure

representing the relationship between exercises to the input data and compressed the input vector to solve the dimensionality problem. Research outcomes proved that the area of the proposed deep KT model was 0.7761, which was better than other KT models and ran faster [14]. This research addressed the dimensionality problem in KT for online teaching and learning systems. Liu S et al. found that the current KT model only attributed the learner's feedback, while ignoring the influence of mental ability factors. Therefore, a new ability-enhanced KT model has been found. It introduced the ability factor into the attribution of feedback, and designed a continuous matrix factorization model. The results proved that the proposed method was better than other KT algorithms in terms of quantitatively evaluated predictive accuracy [15]. This study innovatively introduced the influence of mental ability factors to optimize the KT system more comprehensively.

To sum up, DL is widely used in computer algorithms to improve the accuracy of algorithm models. At the same time, KT models have also been widely used and developed in online education. Therefore, this study innovatively integrates DL and KT, and proposes a new type of traditional cultural online education system. To further improve the interpretability of the deep KT model, the study is based on a dynamic key-value memory neural network, which incorporates two main memory modules for storing knowledge conceptual information and students' state information. In addition, the study recodes the data with the embedding method to avoid the waste of memory resources caused by too high dimensionality. The study aims to provide a more efficient method for accurate recommendation of online learning resources.

3. DL combined with KT online education model.

3.1. Improvement of KT algorithm model based on DL. With the popularity of online education, KT has become one of the best hot research directions. KT uses students' historical learning information to mine students' learning data for online education, with the purpose of improving the correct rate of predicting students' correct answers to questions [16]. The traditional DL-based KT algorithm model, which is named deep knowledge tracing (DKT), is constructed on the long-short-term memory network (LSTM) model. The LSTM model is based on the recurrent neural network (RNN), adding forgetting, input and output gates, so that the network can remember a longer input process [17]. The DKT model is shown in Figure 3.1. The input data type of the model is a $0, 1^{2N}$ sequence, in which N is the number of items. Only one position in the sequence is 1, and the others are 0. For example, if the record "the answer to the first a question is correct", the first element a in the code is 1, and the others are 0. If it is recorded that "the answer to the first a question is wrong", the first element $N + a$ is 1, and the others are 0. The model is an encoder-decoder type model whose output is a sequence of length. During training, the one-hot code $t + 1$ at $q^t + 1$ moment contains both the topic information and the correct and incorrect information. The topic concept scalar at moment corresponds to the prediction information of the next moment in the output sequence. The input gate formula is shown in equation 3.1.

$$i_t = \sigma(W_i[h_t - 1, x_i]) + \sigma b_i \quad (3.1)$$

In equation 3.1, i_t represents the input gate at t moment; W_i means the weight matrix of the input gate; $[h_{t-1}, x_t]$ denotes the connection of two vectors into a longer vector; h_{t-1} and x_t expresses the output items in the two matrices, respectively; b_i refers to the bias of the input gate; the set item *sigma* means the Sigmoid function. The formula of the forget gate is shown in equation 3.2.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t]) + \sigma(b_f) \quad (3.2)$$

In equation 3.2, f_t stands for the forget gate at t time; W_f refers to the weight matrix of the forget gate; b_f represents the bias term of the forget gate; σ expresses the Sigmoid function. The forget gate memory formula is shown in equation 3.3.

$$\widetilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t]) + \tanh(b_c) \quad (3.3)$$

In equation 3.3, \widetilde{C}_t refers to the forget gate memory at t moment; W_c denotes the weight matrix of the forget gate memory; b_c indicates the bias item of the forget gate memory; \tanh means the hyperbolic tangent function. The output gate formula is shown in equation 3.4.

$$O_t = \sigma(W_o \cdot [h_{t-1}, x_t]) + \sigma(b_o) \quad (3.4)$$

In equation 3.4, O_t indicates the output gate at t moment; W_o is the weight matrix of the output gate; b_o stands for the bias item of the output gate; σ expresses the Sigmoid function. Long memory and short memory are shown in equation 3.5 .

$$\begin{cases} h_t = o_t * \tanh(C_t) \\ C_t = f_t * C_{t-1} + i * \widetilde{C}_t \end{cases} \quad (3.5)$$

In equation 3.5, h_t and C_t denote long memory and short memory, respectively. f_t means forget gate, \widetilde{C}_t denote forget gate memory; O_t indicates output gate; \tanh refers to hyperbolic tangent function. The above formula shows that the output in the model is determined by the long memory, and the long memory is affected by the short memory, so the output is affected by the current input and the long-term input. In the LSTM model, long and short memory vectors are passed between each unit [18]. The improved model of this study is based on the dynamic key value memory neural network with two main memory modules. The memory module composed of M and d vectors stores the knowledge concept information and the student's state information, respectively. M is a parameter that can be set, and the dimensions in the module d can also be different. Another improvement is to re-encode the data using the embedding method. Through the embedding matrix, the length of the data $2N * d$ is reduced to the length of the data $2N$, avoiding the waste of memory resources caused by excessive dimensionality. The improved model is shown in Figure 3.2.

As shown in Figure 3.2, the improved algorithm model mainly consists of the embedding process, the graph process, and the knowledge concept encoding process, which correspond to the blue, green, and yellow labels, respectively. In the figure, the embedding process converts high-dimensional, sparse data (e.g., words, student IDs, topic IDs, etc.) into low-dimensional, continuous vector representations, so that data with similarities are closer to each other in this low-dimensional space. In the education domain, the embedding of students and topics captures their implicit attributes or properties. The purpose of the graph process is to construct a network of relationships between knowledge points or between topics to capture and convey information through graph structures. In each iteration of the DKVMN model, the representation of each node is updated according to its neighboring nodes. In this way, the relationships and hierarchies between knowledge points can be explicitly encoded in the model. The purpose of the knowledge concept encoding process is to encode each knowledge point or concept as a continuous vector that captures the dependencies and complexity between knowledge points. The implementation of knowledge concept encoding first obtains an initial representation of each knowledge point through the embedding process. Then, these representations are refined and optimized through the graph process. Eventually, each knowledge point will have a vector representation that adequately captures the essence of the knowledge and its relationship with other knowledge. The calculation of the weight is processed by the embedding matrix, and the dot product operation is performed on the input data with

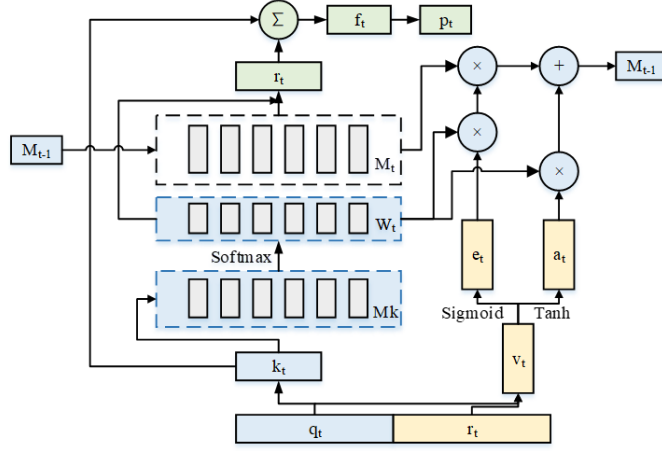


Fig. 3.2: DKVMN model

only the concept of the problem and the vector in the knowledge memory matrix. A vector whose length is equal to the number of vectors in the memory matrix is got, and the weight vector after going through the Softmax layer is got. Softmax is used as an activation function, and its calculation process expression is shown in equation 3.6.

$$w_t(i) = \text{Softmax}(k_t^T M^k(i)) \quad (3.6)$$

As shown in the formula, the activation function Softmax will output the sum of the weight vector elements equal to 1. The improved algorithm model uses knowledge memory and state memory modules. The state memory module is used to record the state at a certain moment, and the weight vector is used to weight and add the vectors in the state matrix during reading M . Embedding matrices are enabled to further correlate inputs in global gradient descent optimization. The reading process is shown in equation 3.7.

$$\begin{cases} r_t = \sum_{i=1}^N w_t(i) M_t^v(i) \\ f_t = \tanh(W_1^T [r_t, k_t] + b_1) \\ p_t = \text{Sigmoid}(W_2^T f_t + b_2) \end{cases} \quad (3.7)$$

In equation 3.7, r indicates the product sum of the weight and the state matrix, which is the value obtained after the connection of r and k ; p indicates the output of the Sigmoid function. The writing part is to adjust the matrix according to the knowledge concept at t moment and the information of whether the answer is correct. The input information generates vectors that control forgetting and memory after embedding. In the writing, both the control memory vector and the previous vector are affected by the association weight, and this improvement makes the process of writing affected by the relation of knowledge concepts. The process of updating the state matrix is shown in equation 3.8.

$$\begin{cases} e_t = \text{sigmoid}(E^T v_t + b_e) \\ \widetilde{M}_t^v(t) = M_{t-1}^v(t) [1 - w_t(i) e_t] \\ a_t = \text{Tanh}(D_{v_t}^T + b_a)^T \\ M_t^v(i) = \widetilde{M}_{i-1}^v(i) + w_t(i) a_t \end{cases} \quad (3.8)$$

In the analysis of the improved model, the study find that the three main modules of the model interact with each other, which will complicate the interaction between knowledge concepts and states, thereby reducing

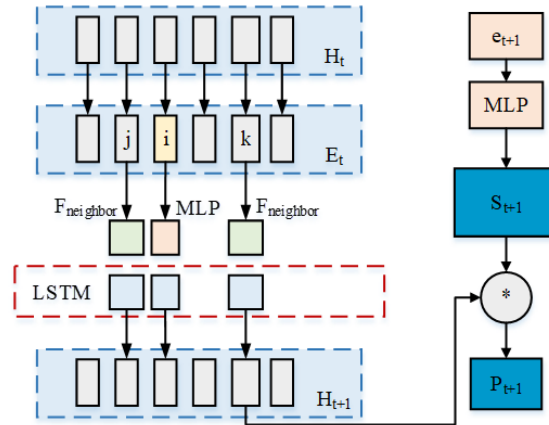


Fig. 3.3: Optimized DL model

the interpretability of knowledge concepts. Therefore, this study introduces deep reinforcement learning for personalized recommendation. The inputs of reinforcement learning include states, actions, and rewards. The key to implementing reinforcement learning is to output the optimal action in the current state to maximize the reward. In reinforcement learning, the long-term rewards obtained through training are optimal, so the summation of rewards is very important. The study introduces a discount factor γ for weighted summation. The expression of weighted summation is shown in equation 3.9.

$$U(S_o) = E\left[\sum_{t=1}^{\infty} \gamma^t R(S_t)\right] \quad (3.9)$$

In equation 3.9, U stands for the sum of rewards, that is, the average expectation of the weighted sum of rewards for all future action choices. R indicates the reward, that is, the system determines the value of an action S . S represents the state, which is the only description of the current environment. This study exploits a recurrence relation when calculating total value to obtain a policy function that maximizes long-term returns. Ideas in reinforcement learning are utilized to divide the long sequence of action decisions, and the mathematical expression is shown in equation 3.10.

$$U^*(S) = \max_a R(a) + U^*(S') \quad (3.10)$$

In equation 3.10, S expresses all the best paths; a denotes the current moment selection action; S' means the remaining paths. After optimization, a small part of the end of the optimally selected path can be truncated, and the remaining path will also become the optimal path.

3.2. Construction of KT online education model based on DL. Some studies have introduced two memory modules to store knowledge status and conceptual relationships, increasing the interpretability of the traditional cultural KT [19]. There are some classic theoretical models in the KT problem, including the combination of dynamic key-value storage network for KT and item response theory of classical cognitive diagnosis model. On this basis, some scholars proposed the knowledge query network (KQN) structure, which increases the interaction between knowledge status and knowledge concepts, to clarify the principle of the specific operation of the model, and further enhance the interpretability of the KT model [20]. The optimized DL model is shown in Figure 3.3. In the Figure 3.3, the improved DL model includes three parts. The first part is the embedding process, in which H is a memory matrix, representing the matrix that stores each knowledge state. H contains N independent vectors, in which N is the total number of problem concepts [21]. H_t

denotes each vector dimension, means the time-memory matrix. E_t indicates the result matrix after inputting embedding processing. The moment matrix E_t is a matrix containing $N - e$ dimensional vectors, and the expression is shown in equation 3.11.

$$E_k^t = \begin{cases} x^t E(k = i) \\ E_c(k)(k \neq i) \end{cases} \quad (3.11)$$

As shown in equation 3.11, the second part is the graph process. The input at t time is (q_t, a_t) . q_t denotes the concept of the problem at t time and the adjacent nodes are recorded as j, k . At this time, only the vectors j and k in the memory matrix are updated. The update process goes through a fully connected part and a LSTM network [22]. The result is recorded as the vector i in the memory matrix, as shown in equation 3.12.

$$H_k^{t+1} \begin{cases} RNN(f_{MLP}(h_k^t))(k_i) \\ RNN(f_{neighbor}(h_i^t))(k \neq i) \end{cases} \quad (3.12)$$

In equation 3.12, the core of this part is the function of k vector. In this study, the setting threshold $f_{neighbor}$ of the function is improved to simplify the model. When the moment is t, k is the title. k and n vectors are not considered, only the adjacent vectors need to be updated. It assumes that the concept input at $t+1$ time is e_{t+1} , after the feed-forward network, the output is a code of length S_{t+1} , the expression is shown in equation 3.13

$$S_{t+1} = \sigma(W_1) \cdot (\sigma(W \circ \cdot e_{t+1} + b_o)) + b_1 \quad (3.13)$$

The proposed model uses DKT network and the deep reinforcement learning network DQN to solve the problem of large memory requirements and long calculation time of the DKT network. It solves the problem that the DQN network cannot further study the relationship between clustering internal concepts with the help of directed graphs relationship problem [23]. In the application of recommendation algorithms, this study innovatively combines knowledge graph (KG) into an online education system that integrates DL and KT, and establishes a complete traditional cultural online education system [24]. KG obtains potential links between items through the description of semantic associations, expands users' interests through various types of associations, improves the accuracy and diversity of recommendations. KG can also maintain users' historical learning data well, increasing interpretability of recommendation results. In the improved recommendation algorithm, the conceptual calculation of the user's operation on the item is shown in equation 3.14.

$$P_i = \frac{\exp(v^T R_i h_i)}{\sum_{(h,t,r) \in S_u^1} \exp(v^T R_i h_i)} \quad (3.14)$$

In equation 3.14, v^T indicates the candidate item vector; R_i mean the edge of the i node; h_i represents the feature vector; S_u^1 refers to propagation process set. The user representation is obtained according to the probability, and the calculation process is shown in equation 3.15.

$$\begin{cases} O_u^1 = \sum_{h_i, r_i, t_i \in S_u^1} p_i t_i \\ u = o_u^1 + o_u^2 + \dots o_u^H \end{cases} \quad (3.15)$$

In equation 3.15, t_i stands for the feature vector of the node's tail node. The similarity between the user and the item is calculated to complete the recommendation at last. The expression is shown in equation 3.16.

$$\widehat{y}_{uv} = \frac{1}{1 + \exp(-u^T v)} \quad (3.16)$$

In equation 3.16, \widehat{y}_{uv} indicates the user u 's click probability for the item v . The research and design of the learning system of traditional cultural network online education divides users into students and teachers

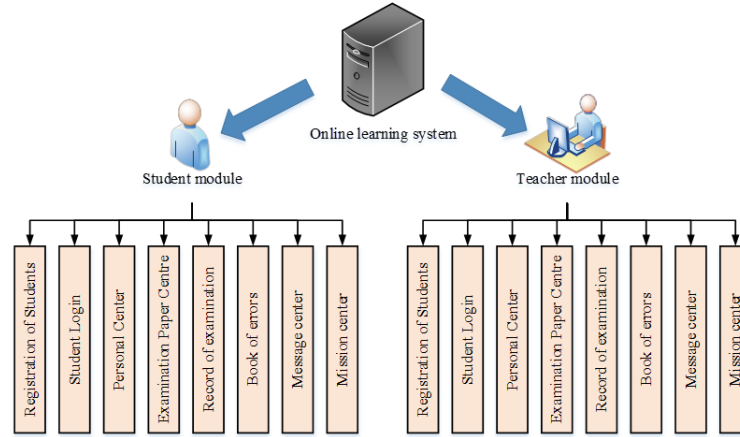


Fig. 3.4: System function overview diagram

Table 3.1: Data set information

Data Set	Algebra_2006_2007	Assistments2009-2010
Number of Logs	52324	58851
Individual Users	511	976
Number of Test Questions	1734	12461
Points of Knowledge	216	110

[25]. Teachers are complex in student management, knowledge management, topic preparation, task release, etc.; students are responsible for completing tasks, taking exams, and answering questions online. An overview of system functions is shown in the Figure 3.4.

In the overall process of traditional cultural online education, the teacher first logs in to the system, checks the students' learning and answering conditions, changes the topics and publishes learning tasks; then the students log in to the system, checks the tasks and notifications, and conducts answering training and knowledge learning. System statistics record students' answers, and input the knowledge structure relationship hidden behind the questions into the model of research design to obtain the sequence of topic recommendations. Finally, students conduct feedback exercises through recommended topics, conduct targeted training, and improve learning efficiency. The overall flow chart of traditional cultural online education is shown in the Figure 3.5.

3.3. Performance comparison results of KT models based on fusion DL. To verify the effect of the knowledge recommendation model based on KT proposed in this study, the experimental data set used algebra_2006_2007 and Assistent2009-2010 data sets that were often used in the field of KT. The algebra_2006_2007 data set was the interaction record between the user and the computer-aided system, and the Assistent2009-2010 data set recorded the practice log of elementary school students' mathematics exercises, including 62955 records. The ratio of the training, test, and verification sets used in the research was 8:1:1. After preprocessing the data, the relevant information of the data is shown in Table 3.1.

The study first explored the impact of different discount factors and different lengths of question sequences on the system, and then established evaluation indicators including area under curve (AUC), accuracy and knowledge mastery fluctuations to verify the improved model's performance proposed in this study. Under different discount factors, the changes in the cumulative rewards of the algorithm model during training iterations

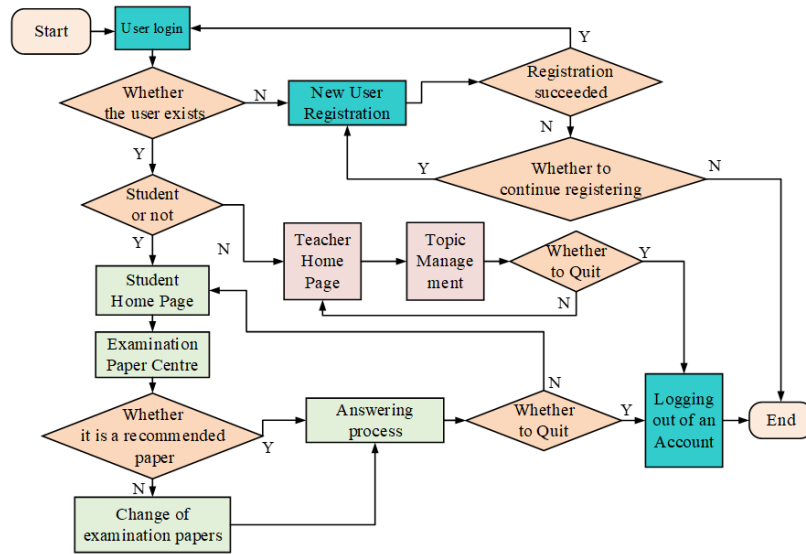


Fig. 3.5: Traditional culture online education flow chart

are shown in Figure 3.6.

From Figure 3.6a, a clear trend can be observed with a discount factor set at 0.4. The cumulative reward of the improved algorithm model exhibited a consistent increase as the number of training iterations escalated. This incremental pattern suggested a potential correlation between the training iterations and the model's efficacy. Conversely, Figure 6(b) portrayed a different scenario when the discount factor was zeroed. The cumulative reward oscillated between 6.32 and 8.47 throughout the training iterations. Notably, for up to 500 training iterations, the reward's trajectory appeared relatively static, devoid of any pronounced pattern. A deeper dive into Figure 3.6 revealed an enlightening insight: the model harnessing a 0.4 discount factor outperformed its counterpart, especially as iterations proliferated. This underpinned the assertion that a long-term recommendation strategy held an advantage over a heuristic approach. To further underscore this, it was essential to test the improved model's capability across various topic sequence lengths. Topic sequences of different lengths were recommended to verify the recommendation ability of the improved model. The result is shown in the Figure 3.7.

Figure 3.7 provided an in-depth analysis of the fusion DL and KT recommendation model across varying application settings. Specifically, Figure 3.7a demonstrated the model's performance when recommending 3 out of 5 questions. By approximately the 150th training step, the model verged on full convergence, maintaining a steady trajectory with an overall reward range between 0.38 and 1.53. Meanwhile, Figure 3.7b, which depicted the recommendation of 4 out of 8 questions, revealed a similar pattern of stability by around the 275th training step, having a reward range of 1.78 to 2.25. However, when ventured into more complex scenarios, the dynamics shifted. Figure 3.7c, representing 5 out of 15 recommendations, suggested that the model struggled to converge even after 500 training iterations, with a reward spectrum of 3.05 to 4.03. Similarly, Figure 3.7d which tackled 10 out of a hefty 50 questions confirmed this challenge. Despite 500 training cycles, convergence remained elusive, and rewards oscillated between 6.26 and 8.55. Drawing insights from these findings, a pattern emerged: the model's convergence and stability were inversely proportional to the number of topics to be recommended. For compact topic sets, the model converged swiftly. However, as the breadth of topics expanded, the convergence rate dwindled, culminating in the model's inability to adequately cater to personalized learning needs across a diverse student body. AUC represents the area enclosed by the ROC and the X coordinate, which can better reflect the pros and cons of the classifier. The closer the AUC value was to 1, the better the classifier effect. Common algorithms KT, DKT, DKVMN, GKT and DL-KT algorithm model proposed in this study were

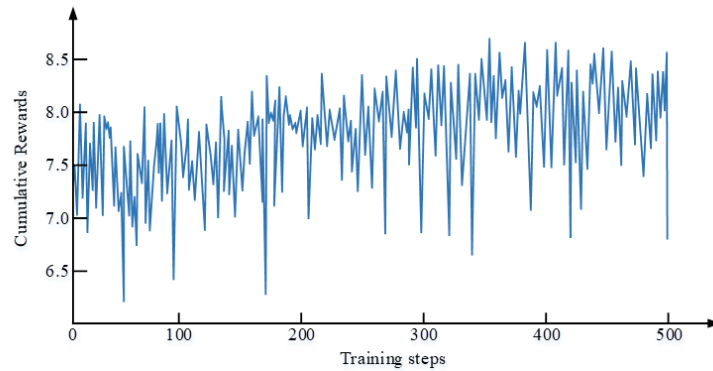
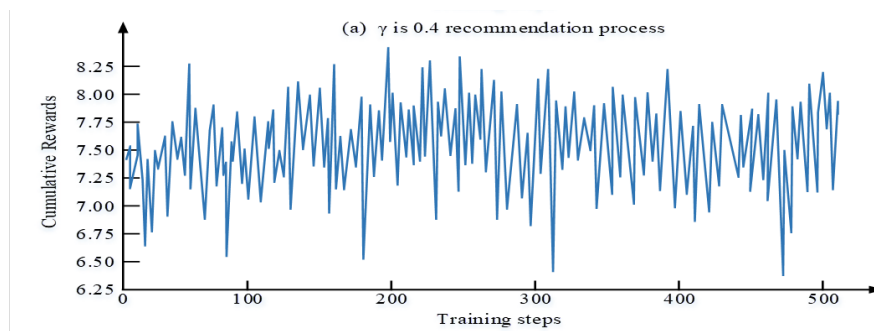
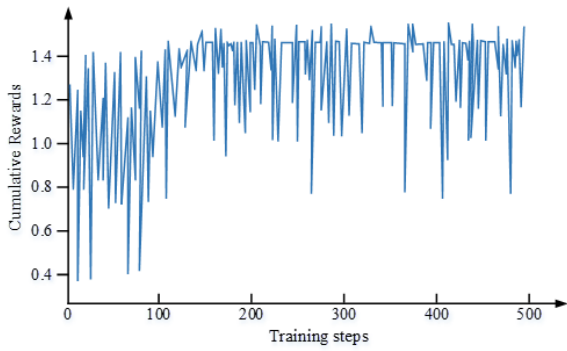
(a) γ is 0.4 recommendation process(b) γ is 0 recommendation process

Fig. 3.6: Training performance of the algorithm model under different discount factors

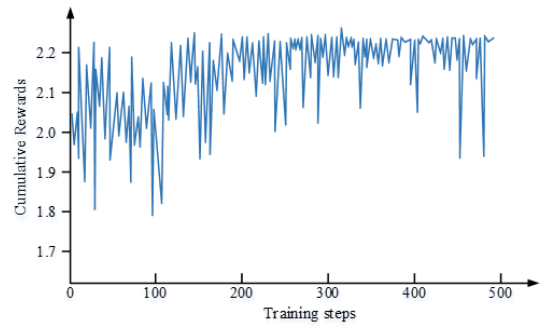
compared. The experimental results of AUC and accuracy are shown in Figure 3.8.

Figure 3.8a shows the AUC comparison of different algorithm models on different training sets. From the figure, the DKT algorithm performed the worst, with an average AUC of 0.762 in the two training sets, followed by the DKVMN, with an average AUC of 0.769, and then the GKT algorithm, with an average AUC of 0.774. The best performance was the algorithm model proposed in this study. The average AUC of the DL-KT algorithm was 0.781. Figure 3.8b shows the comparison of the accuracy of different algorithm models in different training sets. From the figure, the DKVMN performed the worst, with an average accuracy rate of 0.846, followed by the DKT algorithm, with an average accuracy rate of 0.857. The GKT algorithm performed better, with an average accuracy rate of 0.863. The best performer DL-KT was with an average accuracy of 0.886. From the comparison in Figure 8 the DL-KT model proposed in this study performed better than other models in AUC and accuracy. It was verified that the construction mode of the DL-KT model that combined the knowledge display relationship with the potential relationship of knowledge points would make the effect of KT better. The fluctuation of knowledge mastery value (KMV) refers to the fluctuation value between the current knowledge mastery and the previous knowledge mastery after completing the recommended knowledge. KMV is the knowledge fluctuation numerical index. A comparative experiment was carried out on the degree of mastery of knowledge points, and the results are shown in the Figure 3.9.

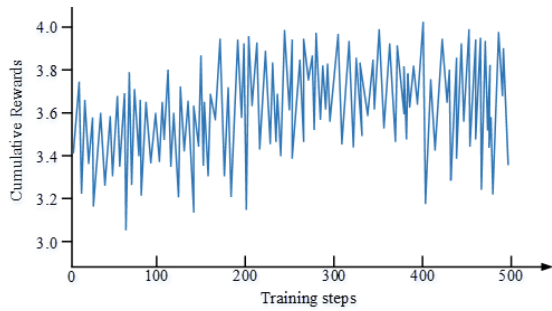
From Figure 3.9, in the Assistent2009-2010 data set, the KMV was the highest when the student's knowledge mastery in the algebra_2006_2007 data set was 0.63. On the whole, the change trend of KMV in the two training sets was the same, and the overall KMV fluctuation range was within 0 to 0.16. The results showed that recommending knowledge within the range of 0.6-0.7 to users improved the user's knowledge



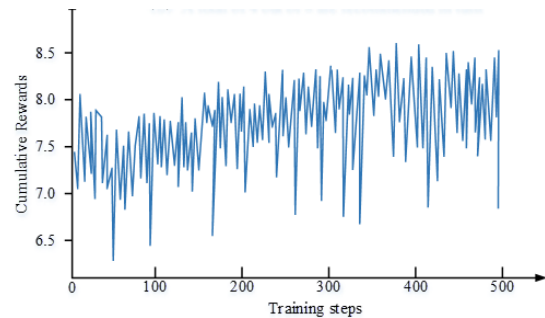
(a) A total of 3 out of 5 are recommended in turn



(b) A total of 4 out of 8 are recommended in turn

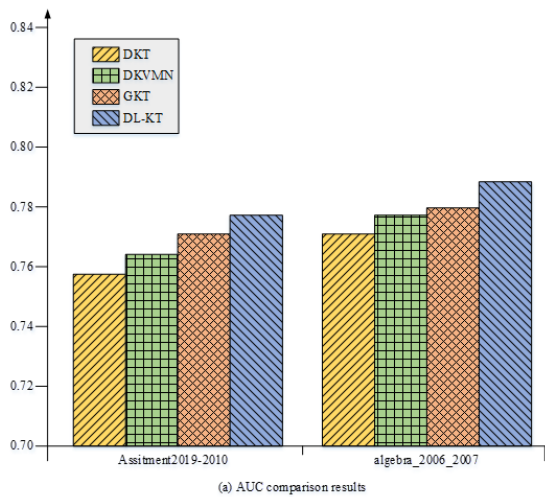


(c) A total of 5 out of 15 are recommended in turn

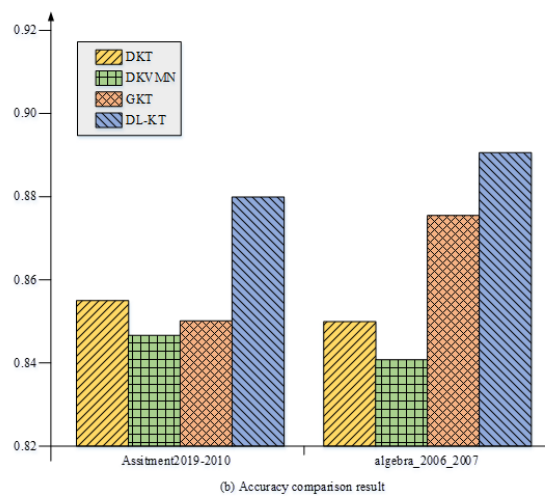


(d) A total of 10 out of 50 are recommended in turn

Fig. 3.7: Model effect performance in different application environments



(a) AUC comparison results



(b) Accuracy comparison result

Fig. 3.8: Comparison of AUC and accuracy of different algorithm models

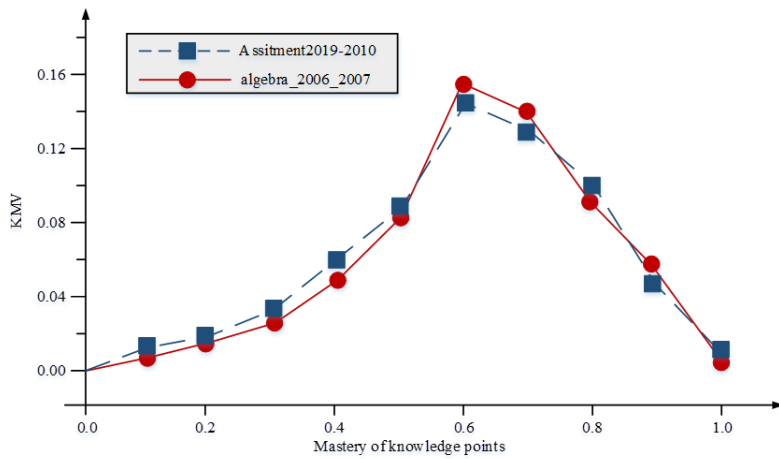


Fig. 3.9: Comparison of knowledge points

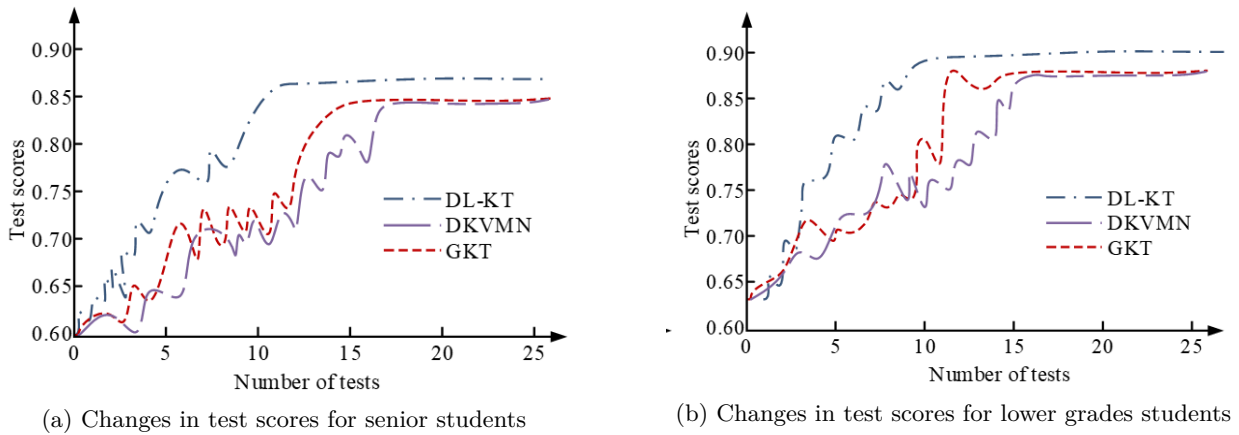


Fig. 3.10: Comparison of application of different models in mental health education

mastery the fastest, which was obviously better than other algorithm models. It was verified the performance of traditional cultural network online education model proposed in this study that integrated DL and KT. To get the effect of the practical application of DL-KT in online education, the study selected senior and junior students to conduct a comparative experiment, respectively, and there was no significant difference in the performance of each group before the experiment. The results of the effect of different models on students' performance in traditional culture online learning were obtained, as shown in Figure3.10.

Figure 3.10a shows the changes in the grades of the students in the higher grades. With the assistance of the DL-KT model, students' grades increased faster and stabilized at about 86 after the 10th quiz. The difference in students' final grades with the assistance of the DINA and HO-DINA models was not significant. Figure 3.10b shows the changes in students' grades in the lower grades. With the assistance of DL-KT model, students' grades increased at a faster rate and their final grades stabilized around 89, which was better than the DINA and HO-DINA models. From Figure 3.10, the DL-KT model proposed in the study was more effective in improving the online learning performance of students in lower grades and was more suitable for application in online education for students in lower grades.

4. Conclusion. The online learning system is a new learning platform that uses the Internet and artificial intelligence technology to enable students to learn the required knowledge more quickly and conveniently. It is difficult to pass the actual level of students in the current online education system and accurately recommend the knowledge question bank. Therefore this research built an online education model that integrated DL and KT. The experimental outcomes denoted that the algorithm model with a discount factor of 0.4 had better performance, which verified that the long-term recommendation effect used in this study was better than the previous heuristic recommendation effect. The average AUC of the model proposed in this study was 0.781, and the average accuracy rate was 0.886. The results showed that recommending knowledge within the range of 0.6-0.7 to users improved the user's knowledge mastery the fastest, which was obviously better than other online education models. However, the shortcoming of this study was that relatively little research has been conducted on the other key subject in online education, the teacher. Teachers play a crucial role in the learning process of students, especially how to track and adapt to the learning progress and needs of different students in real time. In addition, although the model outperformed other models in some metrics, whether it can be generalized across different educational cultures and curricula still requires further research. In future work, the study plans to delve into the role and function of teachers in online learning systems, especially how to enable real-time tracking of student learning by teachers and how to better integrate teachers and technology to provide personalized learning advice to students. As the fields of technology and education converge further, the study hopes to bring about more far-reaching and broader implications for online learning systems.

Fundings. The research is supported by Scientific Research Program Funded by Education Department of Shaanxi Provincial Government (Program No.22JZ016).

REFERENCES

- [1] Dooly, M. & Sadler, R. If you don. *T Improve, What's The Point?' Investigating The Impact Of A 'flipped' Online Exchange In Teacher Education.* **32**, 4-24 (2020)
- [2] Adiego, J. & Martín-Cruz, N. Training competences in smart cities: an online program for higher education students. *International Journal Of Sustainability In Higher Education.* **22**, 1630-1645 (2021)
- [3] Griffith, P., Doherty, C., Smeltzer, S. & Mariani, B. Education initiatives in cognitive debiasing to improve diagnostic accuracy in student providers: A scoping review. *Journal Of The American Association Of Nurse Practitioners.* **33**, 862-871 (2021)
- [4] Root, W. & Rehfeldt, R. Towards a Modern-Day Teaching Machine: The Synthesis of Programmed Instruction and Online Education. *The Psychological Record.* **71**, 85-94 (2021)
- [5] Chowdhury, T., Hoque, M., Wanke, P., Raihan, M. & Azad, M. Antecedents of Perceived Service Quality of Online Education During a Pandemic: Configuration Analysis Based on Fuzzy-Set Qualitative Comparative Analysis. *Evaluation Review.* **46**, 235-265 (2022)
- [6] Shen, L., Shao, Z., Yu, Y. & Chen, X. Hybrid Approach Combining Modified Gravity Model and Deep Learning for Short-Term Forecasting of Metro Transit Passenger Flows: Transportation Research Record. (2021)
- [7] Garg, V. & Wickramaratne, T. Deep Learning Approach for Enhancing Situational Awareness in Surveillance Applications with Ubiquitous High-Dimensional Sensing. *IEEE Journal Of Selected Topics In Signal Processing.* **16**, 869-878 (2022)
- [8] Tang, W., Yang, Q., Hu, X. & Yan, W. Deep learning-based linear defects detection system for large-scale photovoltaic plants based on an edge-cloud computing infrastructure. *Solar Energy.* **231**, 527-535 (2022)
- [9] Bernardini, M., Feldmann, R., Anglés-Alcázar, D., Boylan-Kolchin, M., Bullock, J., Mayer, L. & Stadel, J. From EMBER to FIRE: predicting high resolution baryon fields from dark matter simulations with deep learning. *Monthly Notices Of The Royal Astronomical Society.* **509**, 1323-1341 (2022)
- [10] Liu, C., Han, Z., Zhang, Z., Nussinov, R. & Cheng, F. network-based deep learning methodology for stratification of tumor mutations. *Bioinformatics.* **37**, 82-88 (2021)
- [11] Song, X., Li, J., Tang, Y., Zhao, T., Chen, Y. & Guan, Z. Joint Graph Convolutional Network based Deep Knowledge Tracing. *Information Sciences.* **580** pp. 510-523 (2021)
- [12] Pavlik, P., Eglinton, L. & Harrell-Williams, L. Constrained Framework for Learner Modeling. *Institute Of Electrical And Electronics Engineers (IEEE).* **14**, 624-639 (2021)
- [13] Gan, W., Sun, Y. & Sun, Y. Knowledge structure enhanced graph representation learning model for attentive knowledge tracing. *International Journal Of Intelligent Systems.* **37**, 2012-2045 (2022)
- [14] Huang, Y. & Cheng, Y. Prediction of Online Judge Practice Passing Rate Based on Knowledge Tracing. 2021. *0.* **38** pp. 003 (0)
- [15] Liu, S., Yu, J., Li, Q., Liang, R., Zhang, Y., Shen, X. & Sun, J. Ability boosted knowledge tracing. *Information Sciences: An International Journal.* **596** pp. 567-587 (2022)
- [16] Pawan, Y., Prakash, K., B., C., S., H. & C., Y. Particle swarm optimization performance improvement using deep learning techniques. *Multimedia Tools And Applications.* **81**, 27949-27968 (2022)

- [17] Lai, Z., Wang, L. & Ling, Q. Recurrent knowledge tracing machine based on the knowledge state of students. *Expert Systems*. **38**, 1-12782 (2021)
- [18] Gan, W., Sun, Y., Peng, X. & Sun, Y. Modeling learner's dynamic knowledge construction procedure and cognitive item difficulty for knowledge tracing. *Applied Intelligence*. **50**, 3894-3912 (2020)
- [19] Prakash, A., Patro, K., Hammad, M., Tadeusiewicz, R. & Pławiak, P. secured biometric authentication system using ECG signal based on deep learning techniques. *Biocybernetics And Biomedical Engineering*. **42**, 1081-1093 (2022)
- [20] Huo, Y., Wong, D., Ni, L., Chao, L. & Zhang, J. Knowledge modeling via contextualized representations for LSTM-based personalized exercise recommendation. *Information Sciences*. **523** pp. 266-278 (2020)
- [21] He, J. & Huang, S. Full-length de novo protein structure determination from cryo-EM maps using deep learning. *Bioinformatics*. **37**, 3480-3490 (2021)
- [22] Rakhshani, H., Idoumghar, L., Ghambari, S., Lepagnet, J. & Brevilliers, M. On the performance of deep learning for numerical optimization: An application to protein structure prediction. *Applied Soft Computing*. **110**, 7596-15 (2021)
- [23] Altaheri, H., Muhammad, G., Alsulaiman, M., Amin, S., Altuwaijri, G., Abdul, W. & Faisal, M. Deep learning techniques for classification of electroencephalogram (EEG) motor imagery (MI) signals: A review. *Neural Computing And Applications*. **35**, 14681-14722 (2023)
- [24] Li, Y., Bao, T., Gao, Z., Shu, X., Zhang, K., Xie, L. & Zhang, Z. new dam structural response estimation paradigm powered by deep learning and transfer learning techniques. *Structural Health Monitoring*. **21**, 770-787 (2022)
- [25] Rastogi, K., Bodani, P. & Sharma, S. Automatic building footprint extraction from very high-resolution imagery using deep learning techniques. *Geocarto International*. **37**, 1501-1513 (2022)

Edited by: Mudasir Mohd

Special issue on: Scalable Computing in Online and Blended Learning Environments: Challenges and Solutions

Received: May 17, 2023

Accepted: Sep 11, 2023