DOI 10.31891/2307-5732-2021-303-6-111-114
УДК 004.85

PAVLO RADIUK
Khmelnytskyi National University
ORCID ID: 0000-0003-3609-112X
e-mail: radiukpavlo@gmail.com
OLEXANDER MAZURETS
Khmelnytskyi National University
ORCID ID: 0000-0002-8900-0650
e-mail: exe.chong@gmail.com
TETIANA SKRYPNYK
Khmelnytskyi National University
ORCID ID: 0000-0002-8531-5348
e-mail: marine_1996@ukr.net
OLEKSANDR MOROZ
Khmelnytskyi National University
ORCID ID: 0000-0003-4853-4573
e-mail: pointplack37@gmail.com

INTELLIGENT DATA ANALYSIS USING ARTIFICIAL NEURAL NETWORKS FOR DECISION MAKING IN THE EDUCATION DOMAIN

Nowadays, applying educational intelligent data analysis (*EIDA*) *seems relevant for improving the educational process based on big data. It implies developing and improving the methods of processing collected data in educational institutions to understand academic issues better. Over the past decades, artificial neural networks* (*ANNs*) *have been recognized as the most prominent techniques for learning analytics. In this work, we systematized the recent scientific literature in EIDA with ANNs. The paper analyzes the applications of ANN to EIDA and discusses the computational issues in the EIDA domain. According to the investigation, most educational data mining tasks are addressed by controlled learning models, such as classification, regression, and time-series prediction. Most in-depth methods used in the EIDA domain are traditional types of ANN. Well-known techniques such as multi-year perceptron and deep long short-term memory networks have been mainly used for classification and prediction tasks within the education sphere. However, the difficulty of interpreting the results produced by ANNs has also been a challenge for intelligent data practitioners in any domain, including education.*

Keywords: intelligent data analysis, data mining, artificial neural networks, deep learning, decision making, learning analytics, education

РАДЮК П. М., МАЗУРЕЦЬ О. В., СКРИПНИК Т. К., МОРОЗ О. В.
Хмельницький національний університет

ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ ІЗ ВИКОРИСТАННЯМ ШТУЧНИХ НЕЙРОННИХ МЕРЕЖ ДЛЯ ПРИЙНЯТТЯ РІШЕНЬ У СФЕРІ ОСВІТИ

На сьогодні застосування методів інтелектуального аналізу даних у сфері онлайн освіти видається актуальним для вдосконалення навчального процесу на основі великих даних. Такий підхід передбачає розроблення та вдосконалення методів оброблення зібраних даних у навчальних закладах, щоб краще зрозуміти академічні проблеми та завдання. За останні десятиліття метод інтелектуального аналізу даних, під назвою штучна нейронна мережа, визнано найкращим для різноманітних завдань аналітики. У цій роботі проведено систематизацію нещодавньої наукової літератури щодо застосування методів інтелектуального аналізу даних зі штучною нейронною мережею в галузі освіти. У роботі аналізуються застосування інформаційних технологій на основі штучних нейронних мереж до розв'язання академічних завдань та обговорюються обчислювальні проблеми в галузі освіти. З 2018 року відбувалося стрімке зростання кількості публікацій із використанням штучних нейронних мереж. Половина розглянутих робіт була опублікована американськими установами, тоді як кожну третю статтю опубліковано в азіатських країнах. Відповідно до результатів аналізу встановлено, що освітні дані найчастіше розв'язуються за допомогою контрольованих навчальних моделей. Водночас у галузі освіти найчастіше трапляються завдання класифікації, регресії та прогнозування часових рядів. У роботі визначено, що традиційні типи штучних нейронних мереж найчастіше використовуються для оброблення та аналізу даних галузі освіти. Багатошаровий перцептрон та нейронна мережа з довгою короткочасною пам'яттю використовуються переважно для завдань класифікації та прогнозування. Втім складність інтерпретування результатів, отриманих за штучною нейронною мережею, може бути значною проблемою для фахівців-аналітиків та розробників програмного забезпечення в галузі освіти.

Ключові слова: інтелектуальний аналіз даних, видобуток даних, штучні нейронні мережі, глибоке навчання, прийняття рішень, освітня аналітика, галузь освіти.

Introduction

Due to the coronavirus pandemic caused by COVID-19, 2020 can be considered the year of online education. With the widespread use of information technology (IT), stakeholders in the education domain were able to collect more valuable data on the educational process. Altogether, the application of educational intelligent data analysis (EIDA) seems relevant for improving the educational process based on big data [1]. This methodology implies developing and improving the methods of processing collected data in educational institutions to understand academic issues better. A recent review [2] on the basic analytical approaches used in educational data mining showed that practitioners had used various statistical techniques to facilitate teaching and learning processes. Furthermore, another extensive study [3] from 2020, which covers roughly 300 relevant publications from 2010 to 2018, found that artificial neural networks (ANNs) have become the most widely used methods in EIDA research and would flourish in the future.

An ANN is a simplified computational model of the human central nervous system (CNS) presented in a labelled oriented graph, the vertices of which are in the so-called layers [4]. These vertices accumulate signals from other units and convert the signal to a nonlinear state. They encode the influence of a given block over any other unit that receives a link from the former. Each connection between units can increase or decrease the activation of the receiving unit, and this effect is represented by the so-called weight associated with the graph edges. It learns to perform a task from data that exemplifies the task being performed [5]. Overall, an ANN learning algorithm is a procedure to adapt the weight of the network so that the network begins to perform a computational task.

Analysis of recent research

Over the past decades, ANNs have been fruitfully applied to clustering, classification, regression, time series forecasting, and visualization tasks [2–4]. Nevertheless, the models developed based on the neural network paradigm are recognized as having a lack of explainability [6]. They often give accurate predictions from data, but their results are often difficult to interpret by humans. Nonetheless, despite the difficulties in understanding, ANNs have an excellent ability to learn to develop internally. For example, ANNs have currently considered state-of-the-art solutions in interdisciplinary fields, such as computer vision (CV), image and video generation (IVG), natural language processing (NLP), and audio processing (AP).

In education, ANNs have been used to resolve strategic tasks such as predicting student performance, grouping students according to their characteristics, implementing personalized recommendations that provide adaptive and customized support for student learning, and curriculum planning [7-9]. Meanwhile, a recent review article notes the ANN should not be considered the standard EIDA method because of the conservatism of the educational community and lack of clear explainability [11]. Other recent studies [12-14] identified three areas in education where the use of ANNs for the EIDA tasks can benefit the most: prediction of students' early dropout, student assessment, and intelligent tutoring system.

Problem statement

In this work, we systematized the recent scientific literature in EIDA with ANNs. The paper analyzes the applications of ANN to EIDA and discusses the computational issues in the EIDA domain that can benefit the most by applying ANNs. Thus, the present work aims to study the research in EIDA, conducted from 2017 to nowadays, and used ANNs for data processing.

To achieve the goal of the work, three following tasks are to be resolved.

1. To investigate if ANNs have become widespread in EIDA? If so, what is the balance between traditional types of ANNs and modern deep learning (DL) approaches?

2. To define the primary educational tasks and research problems in EIDA that are addressed with ANNs.

3. To define the ANNs that are state-of-the-art in modern EIDA.

Methodology

The following scientific databases were used for the study: Web of Science Core Collection, Google Scholar, and Scopus. Only works in English are considered. To search for the latest target studies, the following keywords were entered in the search bar: "neural network" or "deep learning," and "educational intelligent data analysis" or "educational data mining." The search period was set from 2017 to 2021. Finally, the study was conducted in October 2021.

The authors of this work do not show the number of excellent general results, as a search in different databases repeatedly returned the same list of works. In addition, four different combinations of subordination, which sometimes had to be used when searching for a given command (for example, "deep learning" and "educational data mining"), returned many repeated results in all searches. Results not related to the purpose of the study were rejected. This hands-on screening was based on the title of each paper, annotation, and keywords. Also, for each of the four different searches performed in Google Scholar, only the first 100 returned results were considered. As a result of this process, 98 articles related to this review were collected.

After the systematization of the found articles, the evaluation of each work was performed based on three criteria. First, only primary studies were included in the analysis, not EIDA surveys or reviews. Second, works without implemented and tested ISs, and without a well-developed theoretical justification were dropped. Third, the target studies were checked for the quality of the submitted materials. In total, eleven articles were excluded from the analysis. As a result of this screening, 87 articles remained to be analyzed to answer three questions formalized above.

Results

Overall, the combined search for the following terms "artificial neural network," "deep learning," "educational intelligent data analysis," and "educational data mining" resulted in 87 relevant publications. The distribution of the found publications on databases is presented in table 1.

Table 1

The number of relevant publications from each scientific base found by searching for key terms

Scientific base	Web of Science Core Collection	Google Scholar	Scopus	Sum
Traditional types of artificial neural networks	6	18	10	34
Modern deep learning techniques	11	27	15	53
Sum	17	45	25	87

Fig. 1 shows the tendency in publications in the EIDA sphere with ANN and DL techniques from 2017 to 2021.

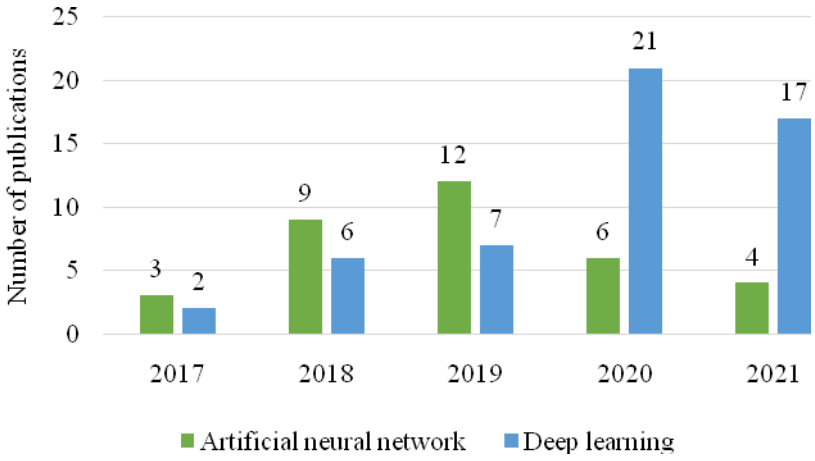

Fig. 1. Number of publications in the EIDA domain with ANN and DL techniques per year from 2017 to 2021

Studies in the EIDA domain appeared before 2017, but according to the explored databases, few works addressed the issue of educational data mining in 2017. However, from 2018 and beyond, the number of publications using ANNs, and DL started to grow significantly. Half of the work was published by American institutions, while every third article was published in Asian countries. No groups have been published twice.

The most prominent studies that address the EIDA tasks with DL techniques are summarized in table 2.

Table 2

Educational intelligent data analysis tasks that are addressed with artificial neural networks and deep learning

Study	Task in education	Description	Technique
[11]	Student assessment	Automizing the processing of students' feedback	Deep long short-term memory network
[7]	Student modeling	Prediction of students' withdraw	Deep long short-term memory network
[8]	Student modeling	Prediction of students' early withdraw	Multi-layer perceptron
[9]	Course video recommendation	Developing a framework for proper recommendation in online learning platforms	Deep long short-term memory network
[10]	Student assessment	Developing an examination generation system	Recurrent neural network
[12]	Student performance prediction	Prediction students' weekly performance by video clickstreams data	Deep long short-term memory network
[13]	Finding an optimal educational facility	Predicting optimal school site selection in the urban areas	Multi-layer perceptron
[14]	Intelligent tutoring system	Developing a virtual patient simulator for clinical diagnostic reasoning	Deep long short-term memory network

Discussion & Conclusion

According to the analysis, it was found that most tasks in educational data mining are addressed by models with controlled learning, such as classification, regression, and time-series prediction. Interestingly, most in-depth methods used in the EIDA domain are traditional types of ANN. Well-known techniques such as multi-year perceptron and deep long short-term memory networks have been primarily utilized for classification and prediction tasks within the education sphere. However, the difficulty of interpreting the results produced by ANNs has also been a challenge for intelligent data practitioners in any domain, including education. Overall, traditional ANNs have achieved impressive results in online education thanks to a significant amount of available data.

References

1. Aldowah H. Educational data mining and learning analytics for 21st century higher education: A review and synthesis / H. Aldowah, H. Al-Samarraie, W. M. Fauzy // Telematics and Informatics. – 2019. – Vol. 37. – P. 13–49.

2. Salloum S. A. Mining in educational data: Review and future directions / S. A. Salloum, M. Alshurideh, A. Elnagar, K. Shaalan // Advances in Intelligent Systems and Computing. – Cham, 2020. – P. 92–102.

3. Charitopoulos A. On the use of soft computing methods in educational data mining and learning analytics research: A review of years 2010-2018 / A. Charitopoulos, M. Rangoussi, D. Koulouriotis // International Journal of Artificial Intelligence in Education. – 2020. – Vol. 30, no. 3. – P. 371–430.

4. Rivas A. Artificial neural network analysis of the academic performance of students in virtual learning environments / A. Rivas [et al.] // Neurocomputing. – 2021. – Vol. 423. – P. 713–720.

5. Радюк П. М. Аналітичний огляд архітектур згорткових нейронних мереж у задачах аналізу медичних зображень / П. М. Радюк // Science, Engineering and Technology: Global and Current Trends : proceedings of inter. and pract. conf. (м. Прага, 27-28 груд. 2019). – Прага, 2019. – С. 32–25.

6. Radiuk P. M. A framework for exploring and modelling neural architecture search methods. / P. M. Radiuk, N. V. Hrypynska // The 4th International Conference on Computational Linguistics and Intelligent Systems (COLINS-2020), Lviv, Ukraine, 23–24 April 2020. – CEUR-WS, 2020. – Vol. 2604. – P. 1060–1074.

7. Hassan S.-UI Virtual learning environment to predict withdrawal by leveraging deep learning / S.-Ul Hassan [et al.] // International Journal of Intelligent Systems. – 2019. – Vol. 34, no. 8. – P. 1935–1952.

8. Mubarak A. A. Prediction of students' early dropout based on their interaction logs in online learning environment / A. A. Mubarak, H. Cao, W. Zhang // Interactive Learning Environments. – 2020. – P. 1–20.

9. Xu W. Course video recommendation with multimodal information in online learning platforms: a deep learning framework / W. Xu, Y. Zhou // British Journal of Educational Technology. – 2020. – Vol. 51, no. 5. – P. 1734–1747.

10. Datta D. Optimization of an automated examination generation system using hybrid recurrent neural network / D. Datta, R. Agarwal, I. Tuteja // International Journal of Interdisciplinary Global Studies. – 2020. – Vol. 14, no. 4. – P. 246–255.

11. Nguyen Ph. X. V. Deep learning versus traditional classifiers on vietnamese students' feedback corpus / Ph. X. V. Nguyen, T. T. T. Hong, K. V. Nguyen, N. L.-T. Nguyen // 2018 5th NAFOSTED Conference on Information and Computer Science (NICS), Ho Chi Minh City, 23–24 November 2018 – [S. l.], 2018. – P. 1–6.

12. Mubarak A. A. Predictive learning analytics using deep learning model in MOOCs' courses videos / A. A. Mubarak, H. Cao, S. A. M. Ahmed // Education and Information Technologies. – 2021. – Vol. 26, no. 1. – P. 371–392.

13. Zaheer N. Optimal school site selection in Urban areas using deep neural networks / N. Zaheer, S.-Ul Hassan, M. Ali, M. Shabbir // Journal of Ambient Intelligence and Humanized Computing. – 2021. – P. 1–15.

14. Furlan R. A natural language processing–based virtual patient simulator and intelligent tutoring system for the clinical diagnostic process: simulator development and case study / R. Furlan // JMIR medical informatics. – 2021. – Vol. 9, no. 4. – P. e24073.

Рецензія/Peer review : 07.12.2021 Надрукована/Printed :30.12.2021