The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/0737-8831.htm

Analyzing legal education mobile learner's behavior using deep learning under social media

Zhen Chen
School of Marxism, Northeastern University, Shenyang, China

Received 27 September 2021
Revised 9 December 2021
20 January 2022
24 February 2022
Accepted 28 February 2022

Abstract

Purpose – Under emerging social media technology, mobile learners' behavior analysis and legality education have important practical significance. The research aims to detect the mobile learning (M-learning) learners' behavior in legality education under the background of the Internet era and improve the learning and teaching effect of online legality education and law popularization.

Design/methodology/approach – This paper proposes a model based on deep learning (DL) fuzzy clustering analysis (FCA), and bidirectional encoder and decoder (ENDEC) of converter model to detect the mobile learners' behaviors in online legality education under the current social media. Then, the effectiveness of the proposed model is tested. The proposed model expects to be applied to multimedia teaching and law popularization activities and provides some theoretical reference and practical value for improving the effectiveness of online teaching.

Findings – The experimental results show that in the learner behavior detection process of M-learning-oriented online legality education, the model's accuracy can reach 99.8%. The response time is shorter than other algorithms. Overall, the application effect of the proposed model and algorithm is good and can be applied in practice.

Research limitations/implications – The research results may lack universality due to the selected research methods. Therefore, researchers are encouraged to test the proposed methods further. In the future, it is necessary to expand the type and scale of text data to improve the accuracy of data detection.

Practical implications – The research results provide a specific theoretical reference and practical significance for improving the learning effect of online M-learning-oriented legality education.

Originality/value – This paper meets the needs of mobile learner behavior analysis based on social media.

Keywords Deep learning model, Legal education system, Learner behavior analysis model, Network education

Paper type Research paper

1. Introduction

Thanks to the booming Internet technology in the recent era, mobile learning (M-learning) has arisen as a well-rehearsed research theme (Grant, 2019). In particular, M-learning provides a time-space-free content-access approach, thereby dominating the traditional multimedia learning and teaching classrooms (Chavoshi and Hamidi, 2019). Nevertheless, the novel approach has not been without downsides (Zhampeissova *et al.*, 2020). To name a few, for some unsupervised learners lacking initiatives, the effect of M-learning can be much poorer. On the other hand, deep learning (DL) has also seen extensive research and applications throughout the century. As a subset of machine learning (ML), DL extracts useful information via hierarchical feature reorganization and learns the feature patterns (Mater and Coote, 2019). To date, DL has well lent itself to such tasks as natural language processing (NLP), image processing (IP), and text classification (TC) (Hayat *et al.*, 2019). Meanwhile, more DL-based methods are being applied to handle learner-generated data sets, data features and text features (Yadav and Vishwakarma, 2020) to better understand learner behavior and learning conditions. Additionally, the combination of DL with other state-of-art technologies

This research was supported by Liaoning Social Science Fund's key project Research on the Implementation Plan of Personalized Ideological and Political Education for College Students in the New Era (L20AKS003).

Library Hi Tech

0737-8831
DOI 10.1108/LHT-10-2021-0355

is getting common (Da'u and Salim, 2020) to solve more complex problems, such as big data analytics (BDA), computer vision and intelligent control (Ghosh *et al.*, 2019).

For a long time, from schools, society and the state, they all have gradually become aware of the importance of legal education. Traditionally, legal education is mainly conferred through teaching or writing, which most students believe to be boring, thus resulting in a relatively poor effect. Accordingly, educators and researchers have explored diversified, lively or exciting methods, such as films, telecourses and WeChat recommendations in recent decades while popularizing the law (Zhang, 2020). Still, available literature methods on legal education have unveiled some drawbacks, especially in school legal education. Specifically, Internet-based online teaching lacks student-teacher interaction in an absolute sense, and the effect of M-learning-oriented legal education is difficult to be guaranteed under a virtual online environment. In the long run, it will impact teenagers' legal awareness and their positive and healthy psychology and values.

Today, legal education has become essential for economic development and social stability, given the ubiquitous Internet and DL technologies giving birth to new economic activities and scientific explorations. Besides, with the vigorous development of social media technology, the number of mobile learners (learners learning through multimedia and the Internet) are catching up rapidly, highlighting the urgency of online education and law popularization more than ever before. This paper's main contribution and innovation are to monitor the behavior characteristics of learners in legal education under the current M-learning environment. Section 1 introduces the Internet and Short Message Service (SMS) research background. Section 2 expounds on the research theories of fuzzy cluster analysis (FCA), convolutional neural network (CNN) and attention mechanism (AM). Section 3 covers the main methods, which studies the bidirectional encoder and decoder (ENDEC) based on the converter, and implements a DL-based model (Habimana *et al.*, 2020) to monitor the learners' behavior in M-learning-oriented legal education under the current social media. Finally, Section 4 draws the research conclusion. The effect of the model test shows that the proposed model is expected to be applied in multimedia teaching and law popularization in the future. The research results have a positive impact on improving the law popularization of Internet-based M-learning-oriented legal education.

2. Recent related work

M-learning refers to the specific learning process without temporal-spatial constraints via mobile computing devices (MCDs). Concretely, M-learning-oriented MCD can effectively present the learning content and provide bidirectional communication between teachers and learners. M-learning technology can be used to publicize and learn legal education, which is practicable for improving Internet-based M-learning-oriented legal education. Hongdao *et al.* (2018) advanced that economic growth (EGC) significantly improved corruption control, whereas legal education played a partial intermediary role in EGC and enhanced corruption prevention. Hamidi and Jahanshaheefard (2019) summarized the elements of M-learning-oriented education information management system (EIMS), based on which the impact was analyzed for students' satisfaction on their expectations and cognition, as well as the dimensions and different problems of M-learning. The results suggested that M-learning had a positive impact on students' satisfaction. Chavoshi and Hamidi (2019) studied the primary acceptance factors of M-learning in Iran higher institutions. They found that social factors also positively impacted perceived usefulness (PU) and perceived ease of use (PEoU). Steinberg *et al.* (2020) believed that with the popularization of MCD, physical education (PE) was facing some challenges, such as more personalized educational mode and expansion of educational content outside of the school gym. Shodipe and Ohanu (2021) investigated college teachers' attitudes, participation and tendency to use M-learning. They used cluster sampling technology to select samples from colleges and universities (CAUs) in three states of the geopolitical region. The data were analyzed using exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). The study found that teachers' perceived ease of use had a significant positive direct relationship with the actual use of ML. Teachers' personality was positively correlated with PEoU, and teachers' mental health was positively correlated with PU. Singh *et al.* (2021) studied the generation of the progressive model and difficulty calculation on mobile applications and graded students according to their abilities. The study aimed to develop a progressive model with difficulty level calibration according to students' abilities. The results showed that with the improvement of digitization and the development of M-learning, the proposed model enabled professors to evaluate students' understanding and test them on the mobile platform. Thus, available literature indicates that with the development of the Internet, mobile learners often master the knowledge and skills better than those receiving a traditional education.

3. Method

3.1 Fuzzy clustering analysis

Figure 1 unfolds the FCA from three aspects:

As shown in Figure 1, there are three specific types of FCA, in which the fuzzy relations-based classification is not suitable for large amounts of data samples.

The membership function (MF) is usually represented by $\mu B(x)$, a measurement of the membership degree of a given object x to the set B in FCA, with a range of [0,1], that is, $0 \leq \mu B(x) \leq 1$. Then, $\mu B(x) = 1$ indicates that x is the full membership of set B. Suppose that a fuzzy set B' contains n objects, $x1, \cdots, xn$, then the fuzzy set can be expressed as:

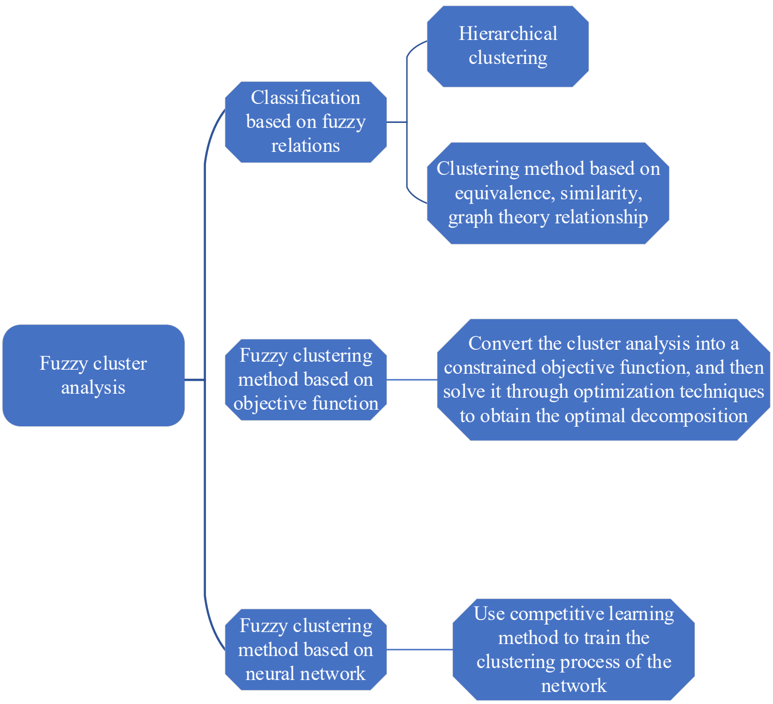

Figure 1.
FCA

$$B' = \{(\mu_B(x_i), x_i) \mid x_i \in X\} \tag{1}$$

Generally, the clustering operation generates a fuzzy set, and the membership degree ranges from [0,1], indicating the degree to which the sample point belongs to the cluster.

Many algorithms can group a dataset in clusters, among which the Hard C-Means (HCM) are most representative. The HCM can divide the given dataset into C-hyper-ellipsoid clusters, which can transform the traditional clustering into nonlinear calculations, as shown in Eq. (2):

$$(\min)J_i = \sum_{i=1}^{C} \sum_{j=1}^{N} u_{ij} \|x_j - v_i\|^2 \tag{2}$$

$U = (uij)C \times N$ represents the HCM clustering matrix, $V = (v1, v2, \cdots, vC)$ is the cluster centers, and $\|-\|$ is the classical Euclidean distance (ED).

The algorithm flow of HCM has three specific steps. (1) It initializes the number of clusters C, the threshold ε for the iteration termination, the original cluster center $V0$ and the iteration count $b = 0$. (2) Step 1: HCM matrix U is updated according to Eq. (3):

$$u_{ij} = \left\{ 1|v_{i-}x_j| \,|= \min 1 \le k \le C\{\|v_k - x_j\|\}0 \ \ \text{others} \right. \tag{3}$$

(3) Step 2: update the cluster center $Vb + 1$;

$$v_i = \frac{\sum_{j=1}^{n} u_{ij} x_j}{\sum_{j=1}^{n} u_{ij}}, 1 \le i \le C \tag{4}$$

(4) Step 3: if $\|Vb - Vb + 1\| < \varepsilon$, the algorithm ends; otherwise, let $b = b + 1$, and it returns to step 1 to continue.

Based on HCM, the clustering algorithm weights the distance between each sample and each cluster center based on the membership square. Then, the objective function (OF) $J1$ of the internal clustering sum of squared errors (SSE) is expanded into the weighted SSE function $J2$, as presented in Eq. (5):

$$(\min)J_2 = \sum_{i=1}^{C} \sum_{j=1}^{N} u_{ij}^2 \|x_j - v_i\|^2 \tag{5}$$

Similarly, the main steps of the FCA process are encapsulated:

(1) Initialize the number of clusters C, the threshold ε, the initial clustering center V and the iteration count $b = 0$.

(2) Step 1: calculate or update the HCM matrix U:

$$u_{ij} = \left[\sum_{k=1}^{C} \left(\frac{\|x_j - v_i\|^2}{\|x_j - v_k\|^2} \right)^{\frac{1}{m-1}} \right]^{-1}, 1 \le i \le C, 1 \le j \le N \tag{6}$$

(3) Step 2: update the cluster center $Vb+1$:

$$v_i = \frac{\sum_{j=1}^{n}(u_{ij})^m x_j}{\sum_{j=1}^{N}(u_{ij})^m}, 1 \le i \le C \tag{7}$$

(4) Step 3: If $||Vb - Vb + 1|| < \varepsilon$, the algorithm ends; otherwise, let $b = b + 1$, and it returns to the first step to continue.

3.2 Convolutional neural network and attention mechanism

For decades, CNN has been under heated discussion, whose architecture includes the feature extractor, generally located in CNN's bottom and middle layers. Meanwhile, CNN also has the convolution layer and the classifier: pooling layer in the bottom of CNN structure, whose output is the input of the fully connected layer. The Softmax function is usually used for multi-class classification by converting input into a probability distribution.

AM redistributes otherwise evenly allocated resources by paying more attention to critical units rather than focusing on unimportant or wrong units. The attention-based resource redistribution is primarily weight assignment in a deep neural network (DNN). Visual attention is divided into several types. The core idea is to find the correlation between them based on the original data and then highlight essential characteristics, including channel attention, pixel attention and multi-order attention. Some also introduce self-AM in the deep residual network (ResNet). AM in neural network (NN) mainly refers to a resource allocation (RA) scheme to alleviate information overload by allocating computing resources to prioritized tasks. Generally, in NN learning, more model parameters often guarantee more robust express ability and more excellent information storage, but complex parameters might cause information overload. By contrast, AM can focus on the prioritized tasks, reduce the attention to sub-prioritized ones and filter out irrelevant information, thereby effectively solving information overload while improving processing efficiency and accuracy.

AM is also divided into many types, as manifested in Figure 2.

As signified in Figure 2, AM includes soft AM, hard AM, self-AM global AM and local AM. Among them, the soft AM calculates the weighted average (WA) of N input information rather than a single information calculation from N pieces of input. In contrast, hard AM selects

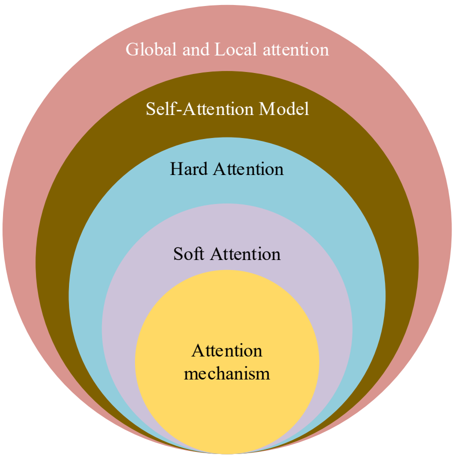

Figure 2.
Classification of AM

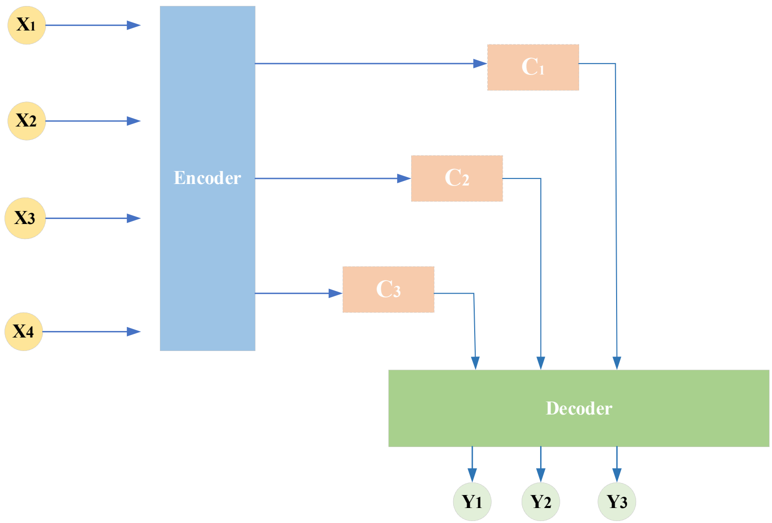

Figure 3.
Structure of ENDEC

information at a specific position in the input sequence, such as random selection or highest probability-based selection. Generally, soft AM is applied in NN; for example, the self-AM model weights different connections between the input and output (not the final output of the model) of the same network layer to obtain the layer-specific network output model. The global AM calculates the output based on all ENDEC input, while the local AM counts the calibration weight based on a few encoders. Figure 3 draws the ENDEC structure (Xiang *et al.*, 2020).

Generally, the AM pattern uses $X = (x1, x2, \cdots, xn)$ to represent N input information. Two steps calculate the attention: (1) The attention distribution is calculated on all input information. (2) The WA of input information is estimated according to the attention distribution. The attention calculation reads:

$$\begin{aligned} \alpha i &= p(z = i|X, q) \\ &= \text{softmax}(s(xi, q)) \\ &= \frac{\exp(s(xi, q))}{\sum_{j=1}^{N} \exp(s(xj, q))} \end{aligned} \tag{8}$$

In Eq. (8), the attention variable $z \in [1, N]$ represents the index position of the selected information. That is, $z = i$ represents the selection of the ith input information. Then, the probability αi of choosing the ith input information is predicted using q and x, where the probability vector composed of αi is called attention distribution. $s(xi, q)$ is the attention scoring function with the following forms:

(1) Additive model:

$$s(xi, q) = v^T \tan h(Wxi + Uq) \tag{9}$$

(2) Dot product model:

$$s(xi, q) = xi^T q \tag{10}$$

(3) Scaled dot product model

$$s(xi, q) = \frac{xi^T q}{\sqrt{d}} \tag{11}$$

(4) Bilinear model:

$$s(xi, q) = xi^T Wq \tag{12}$$

W, U and v represent network parameters for training, and d is the dimension of input information.

Figure 4 displays the attention distribution calculation.

The attention distribution αi represents the correlation between the ith information in the input information vector X and the given query q. The "soft" information selection mechanism is used to query the results. That is, the WA method can summarize the input information to obtain the attention:

$$Att(X, q) = \sum_{i=1}^{N} \alpha iXi \tag{13}$$

ResNet is named after the residual algorithm. Its general structure is outlined in Figure 5 (Wen *et al.*, 2020).

ResNet mainly comprises residual blocks and is designed for a deeper NN structure. Traditionally, as the network layer multiplies, the network degrades. In other words, the loss of the training set first decreases, then saturate, and, finally, increases with the continuous increase of network layers. After the residual block is introduced, a deeper network structure with much better performance can be devised (Chen and Xiu, 2021). The structure of the residual block is demonstrated in Figure 6. Under the general network structure, the input $x1$ directly passes

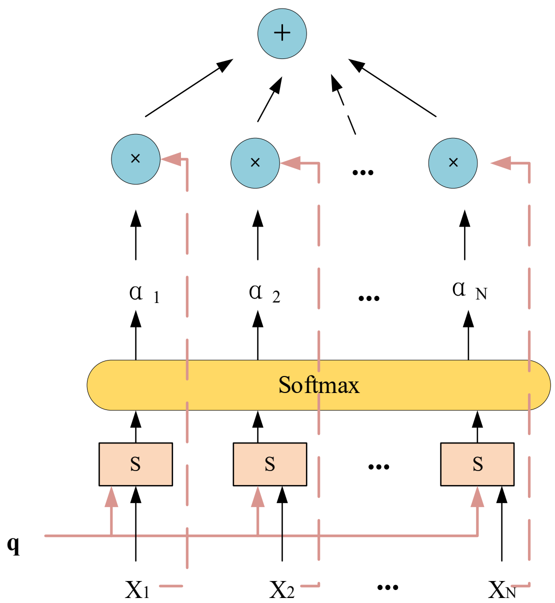

Figure 4.
Calculation of attention distribution

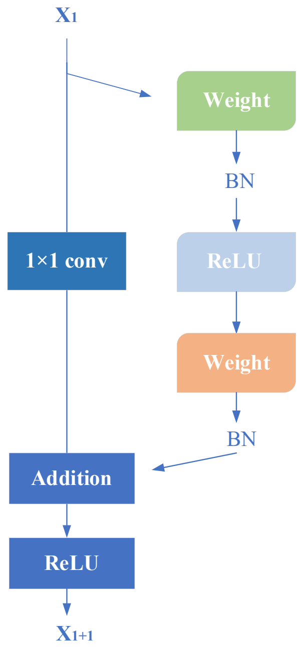

Figure 5.
ResNet structure

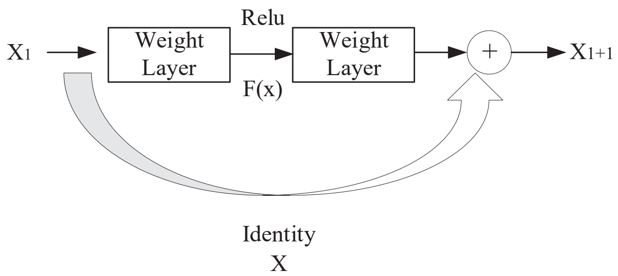

Figure 6.
Residual block structure

through two convolution layers to obtain the output $x1 + 1$. In contrast, the residual block adds the output obtained through two convolution layers to the input $x1$, as confirmed in Figure 6.

The residual function is calculated by Eq. (14).

$$y = F(x,\ W) + x \tag{14}$$

x, W and y denote the input, weight and output of the residual block, respectively.

In the Internet era, the styles of learning, living and entertainment are in a state of flux. Hence, there is a need to detect and analyze learners' online behaviors, such as text, images and video browsing in Internet-based legal education, thereby understanding the learner's specific learning situation. In particular, images are believed to reflect users' static state. For example, by detecting and analyzing the character's clothes or facial expressions in the picture, businesses can capture users' dressing preferences and the instantaneous mood or state. Therefore, ResNet can be used to analyze image-based user learning behaviors.

Facial expression is probably the best form of conveying human emotions and thoughts besides language. Thus, facial expression recognition (FER) can judge the instantaneous state and predict future behaviors of users, and it needs feature extraction (FE). At present, the commonly used FE techniques include local binary pattern (LBP) and principal component analysis (PCA). Researchers have proposed a new FER network training method based on static images (FaceNet2ExpNet) (Hossain and Muhammad, 2019), in which FE is carried out based on $layer2 - layer4$ Dual Attention Feature Enhancement (DAFE) in the ResNet network (Chen *et al.*, 2019). First, CNN is used for multi-scale FE of input facial expressions and image details. Then, the feature map is restored to a specific size by deconvolution. Nevertheless, direct deconvolution often results in feature loss in spatial details and feature point prediction; thus, the DAFE module is introduced to compensate for these losses (Shajini and Ramanan, 2021). FE process is as follows:

$$Fo = D2(D1(F4, F3)F2) \tag{15}$$

$D1$ and $D2$ respectively represent two DAFE modules, and F_2, F_3 and F_4 respectively indicate the $layer2 - layer4$ feature map of ResNet-50. $layer2 - layer4$ of ResNet-50 is used and combined with the DAFE module to realize multi-level feature utilization in the FE structure.

3.3 Text feature representation

This section uses computer network algorithms to deal with the text contents to analyze learners' behavior in legal education using DL. First, the language and symbols in the text must be digitalized. Earlier, some basic language units were mainly regarded as one-dimensional (1D) in vector space and are represented by one-of-v, where v is the size of the dictionary. Later, frequency and information entropy (IE) are introduced into language representation to form typical matrix-based representation methods, such as the latent semantic analysis (LSA) model (Kim *et al.*, 2020). Still, the LSA model cannot express the similarity between words or context relationships and is prone to high omission.

Not until the 21st century did Vaswani *et al.* propose a new network transformer model (Mozafari *et al.*, 2019) according to the AM and apply it in practice. It calculates the input and output representation completely using AM, and is prone to a high omission. The structure of the transformer model is explained in Figure 7.

As illustrated in Figure 7, the transformer model has two encoders. Each comprises six similar modules, includes multi-head AM layer (MHA) and feed forward network (FFN), and is processed by residual and normalization mechanism. MHA is the most crucial part of the model, and its specific structure is revealed in Figure 8 (Yun *et al.*, 2019).

Figure 8 portrays the MHA mechanism to obtain the query's eigenvector, key and value in different subspaces. First, the MHA mechanism radiates them h times linearly, then extracts the triples' features (Q, K, V) from each mapping through the scaling dot product, and finally, splices the attention from H times scaling dot product. The linear layer can fuse the features in different subspaces to get the final output. The attention structure of the scaling dot product is presented in Figure 9 (Yi *et al.*, 2019).

As described in Figure 9, the dot product AM first calculates the point multiplication of query and all keys, divides each by the square root (SQR) $\sqrt{dk}$ to prevent overlarge product, and then feeds it to the Softmax function to obtain the corresponding weight (Jap *et al.*, 2021). According to this weight, it can configure the value vector to get the final output. In Figure 9, the masking process of Q and K are optional. In other words, there is no need to use the encoder to limit the attention module's concern on sequence information. Nevertheless, it is necessary to use the decoder to restrict the attention module only to notice the information of the current time step and the previous time step. This process can be expressed concisely as the attention function (Q, K, V). The attention function (Q, K, V) can calculate the attention

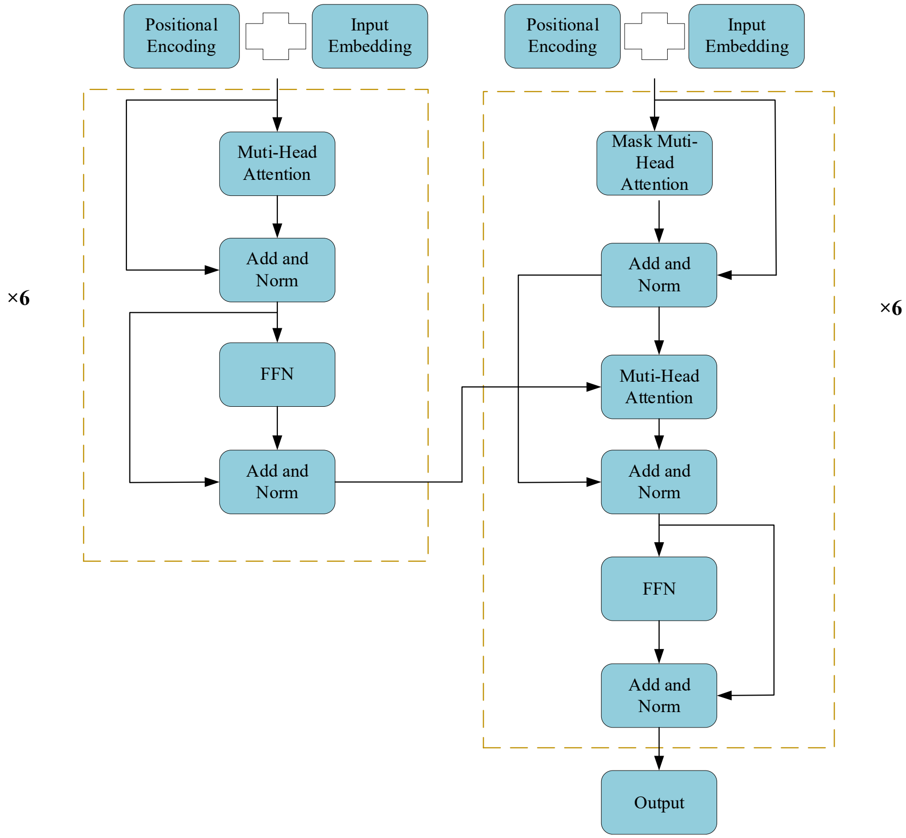

Figure 7.
Transformer model structure

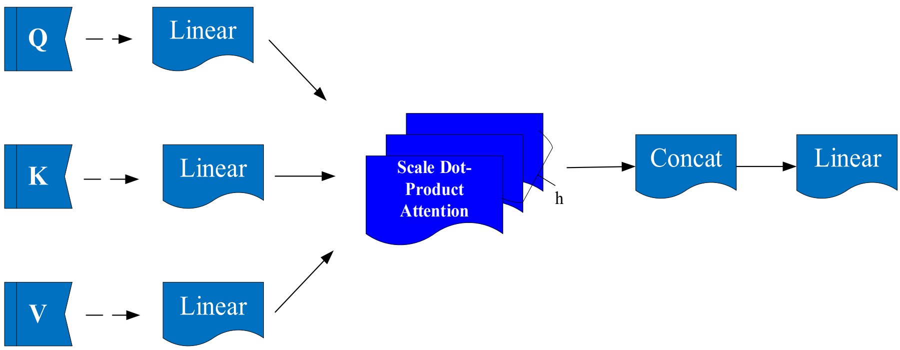

Figure 8.
MHA mechanism

relationship between the query and value sequences when matrices Q, K and V are input. The dimension of Q is $n \times dk$, which means that there are n queries with dimension dk, dimension of K is $m \times dk$ and the dimension of V is $m \times dv$. The product of these three matrices generates an $n \times dv$ dimensional matrix, which represents the value vector noted corresponding to n queries. The specific calculation process is expressed by Eqs. (16) and (17):

$$\text{head } i = \text{Attention } \left(QW_i^o, KW_i^K, VW_i^V\right) \tag{16}$$

$$\text{Multihead}(Q, K, V) = [\text{ head}_1, \text{head}_2, \cdots, \text{head}_n]W^o \tag{17}$$

In Eqs. (16)–(17), Attention$(\cdot)$ represents the AM of the scaling dot product. It first calculates the click similarity of the input query and key. Then, it scales over the input dimension, normalizes and calculates each object's weight in V through the softmax function, and finally uses the point multiplication operation to weighted sum V, as expressed in Eq. (18).

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{18}$$

In particular, this section studies the self-coding language model (LM) structure, which is obtained by bidirectional coding and pre-training target.

3.4 Learners' clickstream FCA of legal education

With the rapid development and application of Internet technology, Internet-based activities are mushrooming. Enterprise servers can trace and record users' clicks, resulting in a tremendous amount of clickstream data. According to these data, learner behavior can be analyzed to uncover their point of interest (POI) or optimize problematic online services. Remarkably, this section defines clickstream as learners' behavior of browsing and clicking the online legal education-related content for learning purposes. Inevitably, the clickstream data are heterogeneous and affect user behaviors from multiple aspects, so it is necessary to use intelligent identification models to analyze learner behavior data. The specific clickstream data can be obtained without changing the original code of the application. The clicking records of learners' browsing behaviors on legal education can be abstracted as:

$$\text{Session} = (P1, T1), \cdots, (Pk, Tk), (\text{null}, Tk + 1) \tag{19}$$

In Eq. (19), Pi represents page attributes, Ti denotes the specific loading time of page Pi, null indicates the end page, and $Tk + 1$ signifies the user-exit time from the page or the software platform. Thus, the page loading time can be changed into the page residence time:

$$\text{Session} = (P1, t1), \cdots, (Pk, tk) \tag{20}$$

In Eq. (19), $ti = Ti + 1 - Ti$, this sequence shows the user's observing behavior process and purpose of learning legal education. The longer the user stays on a page, the higher his interest in the page is. Next, the click records of learners are divided into different clusters and combined with the classical clustering algorithm: density-based spatial clustering of applications with noise (DBSCAN) (Luchi *et al.*, 2019). The specific algorithmic process is displayed in Figure 10 (a)–(c).

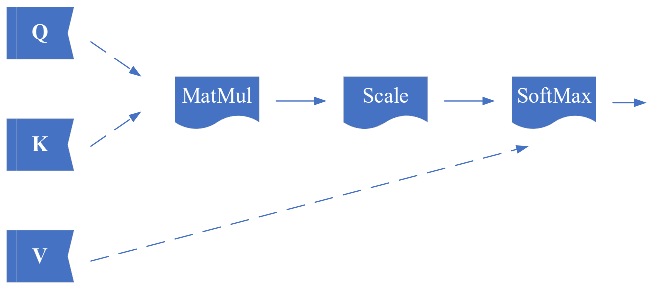

Figure 9.
Dot product AM

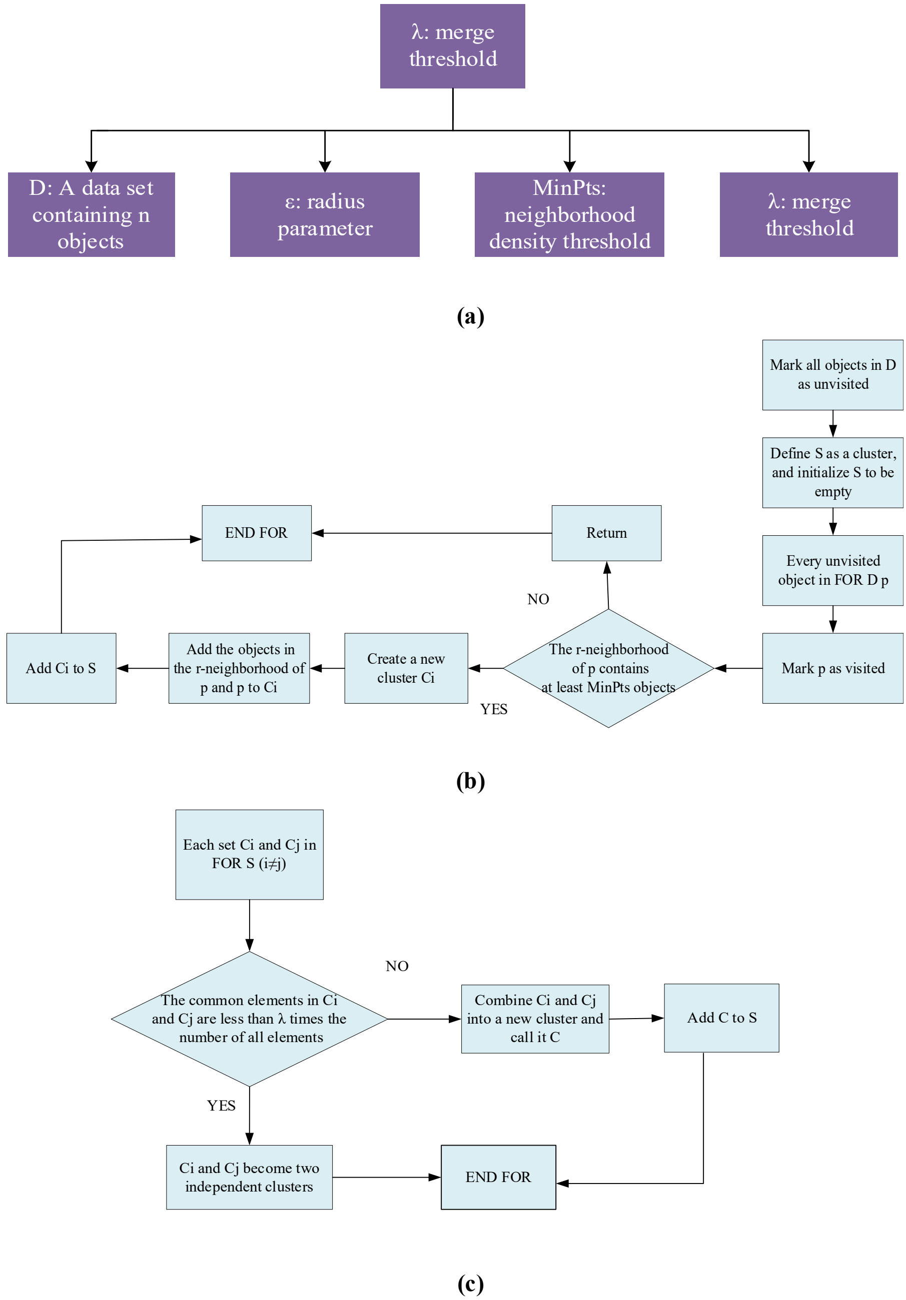

Figure 10.
DBSCAN algorithm process

As reflected in Figure 10 (a)–(c), the DBSCAN algorithm first performs an input process (a), then performs two output processes (b) and (c) simultaneously, and finally outputs the *S* value.

3.5 BERT-based TC of learners' behavior for legal education

Clickstream data in legal education can only give the specific learner behavior and duration but cannot understand the browsed content and learners' POI. Accordingly, this section further analyzes the learner-browsed text information based on clickstream data. Then, a TC model: BERT4TCM-S, is proposed, as sketched in Figure 11 (Gao *et al.*, 2019).

Figure 11 divides the BERT4TCM-S TC model into three parts: the input layer, BERT layer and output layer (Stickland and Murray, 2019). Specifically, the input layer constructs auxiliary sentences according to the category information of different sentence types in the data, transforms the sentence type of the target object into the relationship between the target sentence and the auxiliary sentence, and then uses them for further processing (Vlad *et al.*, 2019). The BERT layer learns the relationship between each unit in the sentence and obtains the semantic feature representation of the input sequence through at least two AM. The output layer predicts the target type according to the semantic representation of the input sequence. This section selects the commonly used topic classification corpus to obtain the statistical information of each corpus to test the performance of the BERT4TCM-S TC model. The experimental data acquisition results are written in Table 1.

According to Table 1, DBPEDIA has ten categories, with each category containing 500 K training data and 60 K testing data. AGNEWS has five categories, with each category containing 100 K training data and 7 K testing data. The DBPEDIA and AGNEWS are regarded as long TC datasets according to their maximum length to 500. In the experiment, the case insensitive pre-training BERT model is uniformly used.

4. Result

4.1 Experimental results of learners' clickstream for legal education

The purpose is to verify the specific application effect of the proposed FCA. Specifically, a network platform is randomly selected, the proposed FCA is applied to the original code of the

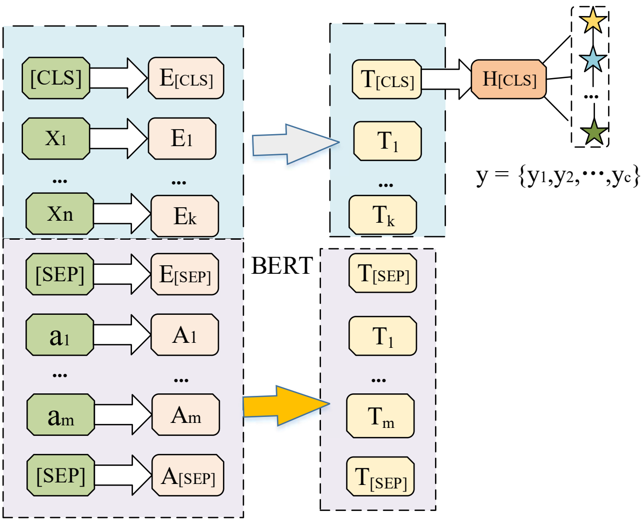

Figure 11.
BERT4TCM-S TC model

platform, and the four sections are tested of social networking, legal education, games and shopping. The distribution of learners' test results is illustrated in Table 2.

Meantime, 25 network platform users are randomly selected and tested. For example, some users are required to answer questions while learning online legal education. Then, 25 learners are tested according to their behaviors, and the results are depicted in Table 3.

As unveiled in Table 3, five users have browsed the legal-related graphics and texts, eight have watched the legal video, seven have answered the legal popularization questions, and seven have searched the legal-related content. Besides, two users have also answered the relevant questions after watching the legal video. Further, Figure 12 compares the improved algorithm with the traditional one.

Figure 12 corroborates that given the same data set, the improved FCA takes less time than the traditional one and, thus, has higher processing efficiency and a more stable clustering effect.

4.2 Experimental results of TC for learners' behavior in legal education

The TC results of BERT4TCM-S with different sentence lengths are outlined in Figure 13.

As Figure 13 suggests, when the maximum length of sentences is 250, sentences shorter than 250 account for 96.5 and 95.4% in AGNEWS and DBPEDIA, respectively. Then, with the increase of the maximum length, this proportion remains the same, so the parameters and calculation of the model will also increase. In particular, the optimal TC effect is reached with the maximum length of 250, which is why the subsequent experiments and practical applications have chosen 250 as the maximum sentence length.

Figure 14 comparatively analyzes the proposed model and other algorithms.

As presented in Figure 14, the accuracy of BERT4TCM-S is better than DPCNN and BERT4TCM-AA. Under the Dbmedia scheme, the prediction accuracy of the proposed BERT4TCM-S can reach 99.8 and 96.7%. The research results prove the particular effectiveness and high accuracy of the improved BERT4TCM-S model.

Table 1.
Basic information of each corpus

Data set	Type	Average length	The maximum length	Training set	Test set	Validation set
DBPEDIA	10	50	3,800	500 k	60 k	10 k
AGNEWS	5	49	1,000	100 k	7 k	10 k

Table 2.
Clickstream test of learners on a network platform

Module content	Social contact	Legal education and learning	Game	Shopping
User attributes	1, 2, 3, 4	3, 4, 5, 6, 7, 8	9, 10, 11, 12, 13	14, 15, 16

Table 3.
Clickstream test for learners of network platform-based legal education

Module content	Browse graphic content	Watch the legal video	Answers to legal questions	Search legal videos
User attributes	1, 2, 3, 4, 5	6, 7, 8, 9, 10, 11, 12, 13	12, 13, 14, 15, 16, 17, 18	19, 20, 21, 22, 23, 24, 25

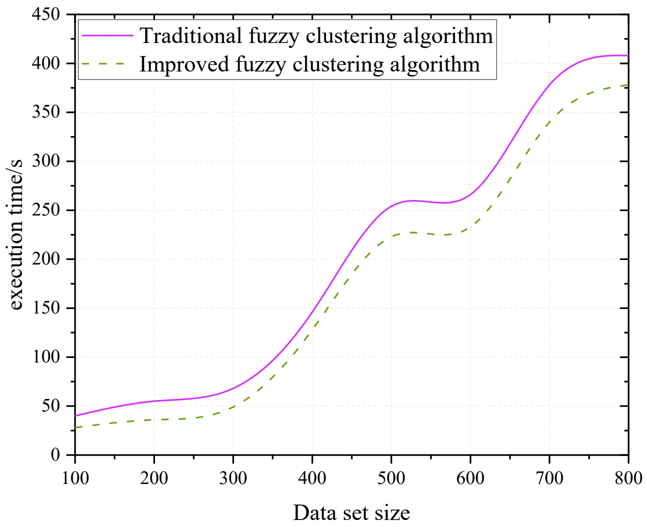

Figure 12.
Data sets processing comparison by different algorithms

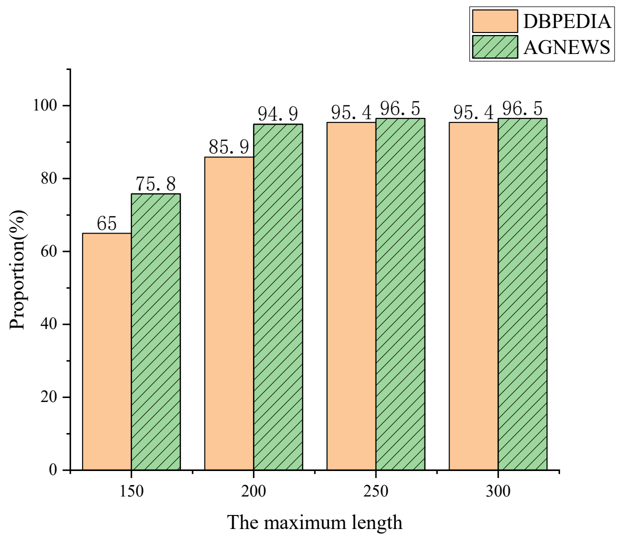

Figure 13.
BERT4TCM-S TC of sentences with different lengths

4.3 Results discussion

With the rapid development of mobile Internet and the advocation of legal popularity, Internet-based social media technologies are seeing more comprehensive applications in multimedia teaching, especially in legal education. M-learning-oriented learner behavior analysis suggests that the legal education-intended online learners account for a large proportion of mobile learners. Most of them watch legal consultation programs or search for legal videos. Meanwhile, with the increase of network users, the mechanism to improve the effectiveness and accuracy of learner behavior recognition demand urgent solutions. Hamidi and Chavoshi (2018) discussed the influence of usability factors on M-learning development

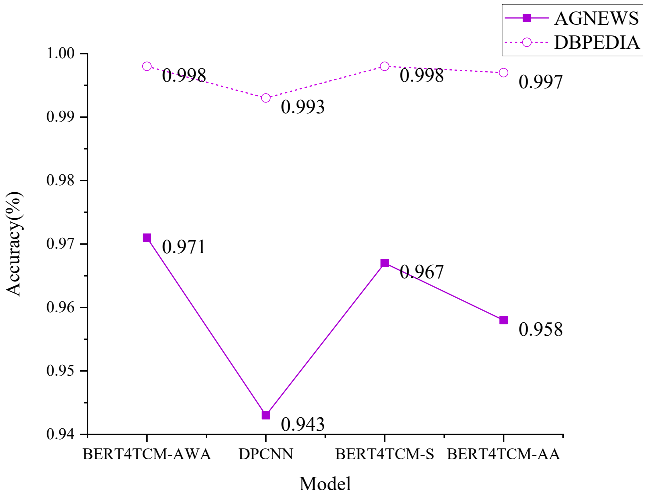

Figure 14.
Accuracy of BERT4TCM-S TC model and other models

in the educational environment and use culture. The results found that personal characteristics and personality factors significantly and positively impacted the use culture, and the culture had a significant positive impact on behavior intention. Pappas *et al.* (2019) used a fuzzy set qualitative comparative analysis for the data samples of mobile learners. They provided new insights on how the predictors of M-learning adoption were interrelated, which had practical reference value for expanding cognitive and emotional characteristics, combining social and personal factors, and highly adopting the existing knowledge of M-learning. Kasakowskij *et al.* (2020) analyzed the increasingly popular antilocution and other violations of personality rights on social media through a critical study of social media-based online law enforcement. The results showed that all countries were strengthening the supervision of online content and creating a cleaner Internet environment. Siregar *et al.* (2020) conducted a law enforcement study on preventing the dissemination of criminal crimes through social media technologies. They defined content containing racial discrimination as antilocution, which needed special and reasonable monitoring. Fithri and Wahyuni (2021) examined the ways of judicial recovery in the crime of social media humiliation. The research results implied that the mode of social harmony and interaction could be stored by prioritizing repairing the initial conditions and dealing with the criminal acts that were fair and commensurate with the victims and perpetrators. Then, the proposed restorative justice approach would be used in the investigation and review stages of the tribunal. Based on the above literature, this paper uses DL and FCA to analyze learners' behavior under social media. By optimizing the TC algorithm, the research improves the accuracy of learner behavior recognition and is also effective for user classification in legal education. Additionally, applying CNN and AM can provide specific theoretical references and practical value for improving Internet-based M-learning-oriented legal education's effectiveness.

5. Conclusion

With the development of mobile information technology, more schools have introduced M-learning and multimedia teaching. However, without supervision, the effect of traditional multimedia teaching is relatively low, mainly because of the lack of autonomous learning.

In this context, this paper proposes a model by fusing FCA and DL to analyze specific learner behavior data, data features and text features in the M-learning process. The experimental results show that the improved FCA takes fewer data processing times than the traditional algorithm. Thus, the proposed algorithm has higher processing efficiency and a more stable clustering effect. Its advantage and novelty lie in using FCA and CNN to analyze learners' behavior data, data features and text features. The proposed model and algorithm have high accuracy in detecting learners' behavior in legal education, which has specific theoretical reference and practical significance for improving the learning effect of M-learning legal education. The research content is generally applicable to supervising learners and improving learning efficiency in M-learning-oriented legal education. However, there are some deficiencies in the research. First, the volume of text data is too small to improve the accuracy of experimental results further. The follow-up research will expand the amount of research text data. Second, the proposed CNN needs data normalization before model training, and the CNN model lacks memory function. It is expected that the type and scale of text data will be expanded to improve the accuracy of learner behavior detection.

References

Chavoshi, A. and Hamidi, H. (2019), "Social, individual, technological and pedagogical factors influencing mobile learning acceptance in higher education: a case from Iran", *Telematics and Informatics*, Vol. 38, pp. 133-165.

Chen, Z. and Xiu, D. (2021), "On generalized residual network for deep learning of unknown dynamical systems", *Journal of Computational Physics*, Vol. 438, p. 110362.

Chen, M., Qin, Y., Qi, L. and Sun, Y. (2019), "Improving fashion landmark detection by dual attention feature enhancement", *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*.

Da'u, A. and Salim, N. (2020), "Recommendation system based on deep learning methods: a systematic review and new directions", *Artificial Intelligence Review*, Vol. 53 No. 4, pp. 2709-2748.

Fithri, B.S. and Wahyuni, W.S. (2021), "Restorative justice approach in crime of humiliation through social media", *Veteran Law Review*, Vol. 4 No. 2, pp. 143-156.

Gao, Z., Feng, A., Song, X. and Wu, X. (2019), "Target-dependent sentiment classification with BERT", *IEEE Access*, Vol. 7, pp. 154290-154299.

Ghosh, S., Das, N., Das, I. and Maulik, U. (2019), "Understanding deep learning techniques for image segmentation", *ACM Computing Surveys (CSUR)*, Vol. 52 No. 4, pp. 1-35.

Grant, M.M. (2019), "Difficulties in defining mobile learning: analysis, design characteristics, and implications", *Educational Technology Research and Development*, Vol. 67 No. 2, pp. 361-388.

Habimana, O., Li, Y., Li, R., Gu, X. and Yu, G. (2020), "Sentiment analysis using deep learning approaches: an overview", *Science China Information Sciences*, Vol. 63 No. 1, pp. 1-36.

Hamidi, H. and Chavoshi, A. (2018), "Analysis of the essential factors for the adoption of mobile learning in higher education: a case study of students of the University of Technology", *Telematics and Informatics*, Vol. 35 No. 4, pp. 1053-1070.

Hamidi, H. and Jahanshaheefard, M. (2019), "Essential factors for the application of education information system using mobile learning: a case study of students of the university of technology", *Telematics and Informatics*, Vol. 38, pp. 207-224.

Hayat, M.K., Daud, A., Alshdadi, A.A., Banjar, A., Abbasi, R.A., Bao, Y. and Dawood, H. (2019), "Towards deep learning prospects: insights for social media analytics", *IEEE Access*, Vol. 7, pp. 36958-36979.

Hongdao, Q., Mumtaz, A. and Mukhtar, H. (2018), "Corruption prevention and economic growth: a mediating effect of rule and law", *International Journal Social Science Studies*, Vol. 6, p. 128.

Hossain, M.S. and Muhammad, G. (2019), "Emotion recognition using deep learning approach from audio–visual emotional big data", *Information Fusion*, Vol. 49, pp. 69-78.

Jap, D., Won, Y.S. and Bhasin, S. (2021), "Fault injection attacks on SoftMax function in deep neural networks", *Proceedings of the 18th ACM International Conference on Computing Frontiers*, pp. 238-240.

Kasakowskij, T., Fürst, J. and Fischer, J. (2020), "Network enforcement as denunciation endorsement? A critical study on legal enforcement in social media", *Telematics and Informatics*, Vol. 46, 101317.

Kim, S., Park, H. and Lee, J. (2020), "Word2vec-based latent semantic analysis (W2V-LSA) for topic modeling: a study on blockchain technology trend analysis", *Expert Systems with Applications*, Vol. 152, p. 113401.

Luchi, D., Rodrigues, A.L. and Varejão, F.M. (2019), "Sampling approaches for applying DBSCAN to large datasets", *Pattern Recognition Letters*, Vol. 117, pp. 90-96.

Mater, A.C. and Coote, M.L. (2019), "Deep learning in chemistry", *Journal of Chemical Information and Modeling*, Vol. 59 No. 6, pp. 2545-2559.

Mozafari, M., Farahbakhsh, R. and Crespi, N. (2019), "A BERT-based transfer learning approach for hate speech detection in online social media", *International Conference on Complex Networks and Their Applications*, Springer, Cham, pp. 928-940.

Pappas, I.O., Giannakos, M.N. and Sampson, D.G. (2019), "Fuzzy set analysis as a means to understand users of 21st-Century learning systems: the case of mobile learning and reflections on learning analytics research", *Computers in Human Behavior*, Vol. 92, pp. 646-659.

Shajini, M. and Ramanan, A. (2021), "An improved landmark-driven and spatial–channel attentive convolutional neural network for fashion clothes classification", *The Visual Computer*, Vol. 37 No. 6, pp. 1517-1526.

Shodipe, T.O. and Ohanu, I.B. (2021), "Electrical/electronics technology education teachers attitude, engagement, and disposition towards actual usage of Mobile learning in higher institutions", *Education and Information Technologies*, Vol. 26 No. 1, pp. 1023-1042.

Singh, R., Timbadia, D. and Kapoor, V. (2021), "Question paper generation through progressive model and difficulty calculation on the Promexa Mobile Application", *Education and Information Technologies*, Vol. 26 No. 4, pp. 4151-4179.

Siregar, G., Siregar, S.A. and Silaban, R. (2020), "Legal implementation of electronic information and transaction law in preventing the spread of content containing SARA issues through social media", *International Journal of Innovation Creativity and Change*, Vol. 13 No. 10, pp. 1418-1431.

Steinberg, C., Zühlke, M. and Bindel, T. (2020), "Aesthetic education revised: a contribution to mobile learning in physical education", *German Journal of Exercise and Sport Research*, Vol. 50 No. 1, pp. 92-101.

Stickland, A.C. and Murray, I. (2019), "Bert and pals: projected attention layers for efficient adaptation in multi-task learning", *International Conference on Machine Learning*, PMLR, pp. 5986-5995.

Vlad, G.A., Tanase, M.A. and Onose, C. (2019), "Sentence-level propaganda detection in news articles with transfer learning and BERT-BiLSTM-capsule model", *Proceedings of the Second Workshop on Natural Language Processing for Internet Freedom: Censorship, Disinformation, and Propaganda*, pp. 148-154.

Wen, L., Li, X. and Gao, L. (2020), "A transfer convolutional neural network for fault diagnosis based on ResNet-50", *Neural Computing and Applications*, Vol. 32 No. 10, pp. 6111-6124.

Xiang, X., Zhang, Y. and El Saddik, A. (2020), "Pavement crack detection network based on pyramid structure and attention mechanism", *IET Image Processing*, Vol. 14 No. 8, pp. 1580-1586.

Yadav, A. and Vishwakarma, D.K. (2020), "Sentiment analysis using deep learning architectures: a review", *Artificial Intelligence Review*, Vol. 53 No. 6, pp. 4335-4385.

Yi, S., Wang, X. and Yamasaki, T. (2019), "Impression prediction of oral presentation using LSTM and dot-product attention mechanism", *2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM)*, IEEE, pp. 242-246.

Yun, S., Jeong, M. and Kim, R. (2019), "Graph transformer networks", *Advances in Neural Information Processing Systems*, Vol. 32, pp. 11983-11993.

Zhampeissova, K., Kosareva, I. and Borisova, U. (2020), "Collaborative Mobile Learning with Smartphones in Higher Education", *International Journal of Interactive Mobile Technologies (iJIM)*, Moscow, Vol. 14 No 21, pp. 4-18.

Zhang, G. (2020), "A study on the application of the film and television works about the law in adult legal education", *Barnard Education Review*, Vol. 1 No. 1, pp. 50-57.

Corresponding author
Zhen Chen can be contacted at: zhenchenneu@163.com