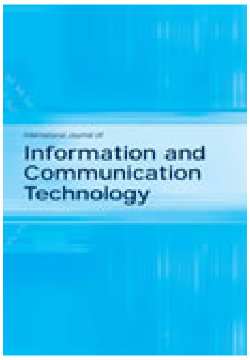


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Evaluation of teaching quality in database courses based on domain-adaptive transfer learning

Xuesong Yang

DOI: 10.1504/IJICT.2025.10071895

Article History:

Received:	06 May 2025
Last revised:	24 May 2025
Accepted:	24 May 2025
Published online:	10 July 2025

Evaluation of teaching quality in database courses based on domain-adaptive transfer learning

Xuesong Yang

College of Electrical Engineering,
Northwest Minzu University,
Gansu 730000, China
Email: yangxuesong2025@163.com

Abstract: The distribution of teaching data varies among database courses, and traditional methods are often difficult to deal with such domain differences effectively. For this reason, this paper firstly utilises BERT model for embedding learning of teaching feedback text, and then extracts local and global features of the text through convolutional neural network (CNN) and long short-term memory (LSTM) network respectively, and enhances the text features through the attention mechanism. On this basis, the domain adaptive transfer learning algorithm is adopted to achieve the characteristic distribution migration alignment of the text source topic and objective topic, and minimise the scoring difference between different classifiers through consistency constraints, so as to assess the teaching quality more accurately. Simulation results show that the classification accuracy of the offered method is 94.39%, which demonstrates a substantial enhancement over the benchmark method.

Keywords: database curriculum; teaching quality evaluation; BERT model; domain adaptation; transfer learning.

Reference to this paper should be made as follows: Yang, X. (2025) 'Evaluation of teaching quality in database courses based on domain-adaptive transfer learning', *Int. J. Information and Communication Technology*, Vol. 26, No. 24, pp.49–64.

Biographical notes: Xuesong Yang received his Master's degree from Universiti Utara Malaysia in 2010. He is currently a Lecturer in the College of Electrical Engineering at Northwest Minzu University. His research interests include information systems and information security.

1 Introduction

Under the impact of the wave of digitalisation, the teaching quality of database courses is not only linked to the cultivation of students' professional skills, but also exerts a profound influence on the promotion of related industries' talent pool and technology development. The traditional teaching quality evaluation system mostly relies on fixed indicators and manual scoring, which has significant deficiencies in the comprehensiveness of data collection, the plurality of evaluation dimensions, and the dynamic adaptability of evaluation models (Ishaq et al., 2023). With the popularity of online education platforms and the massive growth of teaching data, how to fully explore the value of data in the teaching process and construct a scientific, accurate and generalisation-capable teaching quality assessment models has emerged as a critical challenge in the field of education (Musti, 2015). As a major research focus in the domain of machine learning, transfer learning (TL) successfully addresses the issue of limited data availability in the objective domain by transferring the knowledge gained from the source domain to the objective context (Lagus et al., 2018). Although TL has shown some potential for application in the field of education, research aimed at evaluating the teaching quality of database courses is still in its infancy (Tsiakmaki et al., 2020). Most of the existing research stays at the application level of traditional machine learning methods and lacks a systematic solution to the problem of domain differences in instructional data.

Li and Yuan (2021) utilised the multiple linear regression method to evaluate the courses based on the computer application students in a school and obtained a high prediction accuracy. Chen and Mokhtar (2023) applied the time series prediction (ARIMA) algorithm to estimate the instructional assessment outcomes for the information management curriculum, but the prediction error was large. Dong (2023) utilised the decision tree (DT) algorithm to predict students' grades by using factors such as students' high school scores, college grades, and gender as indicators and training data based on a multilayer perceptron topology model. Wang et al. (2022a) combed the relevant indicators affecting course evaluation, and extracted the main indicators through the principal component analysis algorithm, and output the evaluation results through support vector machine (SVM). Hou (2021) proposed a course evaluation model for course performance mining based on SVM and DT, which provided suggestions for instructors' teaching style and course scheduling sequence. Li and Zhang (2024) used clustering algorithm and a priori principle algorithm (apriori) to analyse teachers' teaching quality evaluation results and mine association rules among course grades, which were used to make early warnings to teachers with bad evaluation results.

There are higher-order interactions between teaching quality and influencing factors, which are difficult to be portrayed by traditional linear models. Deep learning establishes implicit associations between multiple factors through multilayer nonlinear transformations to improve the evaluation effect. Liu (2022) used a CNN to capture characteristics from students' evaluation texts and output the prediction results through a fully connected network with high classification accuracy. Sebbaq and El Faddouli (2022) used the BERT model to generate dynamic vectors of evaluation texts, and then the vectors were fed into a gated recurrent unit (GRU) used to categorise the quality of teaching and learning with a classification accuracy of 81.9%. Mao et al. (2024) adopted CNN to capture local characteristics of the evaluated text, extracted global features through LSTM, and fused local and global features through the attention mechanism to improve the classification accuracy.

To address the issue of scarcity of teaching evaluation data samples and labels, TL can learn domain-invariant feature knowledge from domains that are richer in sample data and labels to enhance the generalisation ability of network models. Yun et al. (2023) designed a migration studying approach with CNN and multicore dynamic distribution, which can adjust the feature weights according to the data distribution in the objective field, and applied it to the evaluation of English courses, which achieved good evaluation results. Li (2022) proposed a hierarchical deep educational domain adaptation migration method based on correlation alignment damage function (CORAL), which can adapt to the data and features of different educational domains. Tsiakmaki et al. (2020) suggested a deep TL approach that effectively bridges the gap between source and target domains by leveraging both expert knowledge and adversarial learning. Wang et al. (2022b) constructed an augmented migration CNN, and applied both the categorisation loss function and the classifier discriminant loss function to realise adaptive matching among the objective field samples and the source field samples. Guo et al. (2024) develop a deep adversarial TL approach in light of Wasserstein distance to learn a shared feature representation by employing adversarial training to bridge the domain gap using adversarial learning to narrow the inter-domain divergence.

Based on the analysis of existing studies, it can be seen that current teaching quality evaluation methods ignore the differences that exist in the distribution of teaching data, resulting in poor classification accuracy of the evaluation model. For this reason, this article proposes an approach for assessing the instructional quality within database courses in light of domain-adaptive transfer studying. First of all, the database course teaching quality evaluation text embedded learning, using the text to assess the different aspects related to teaching quality, for the feedback text using BERT-based sentence embedding learning, and combined with the subject lexicon embedding, to lay the foundation for the construction of the subsequent evaluation model. Then the local features of the feedback text are extracted by CNN and the global features of the feedback text are extracted by LSTM, and the text features are enhanced by the attention mechanism. On this basis, in order to study the shared characteristics of each source-topic and objective-topic text pair, this paper generates pseudo-labels for unlabeled target-topic texts by aligning their distributions at the topic level and category level through domain adaptive transfer learning. By aggregating the neighbourhood information to seize the intrinsic data organisation at the local scale, we can minimise the scoring differences between different classifiers and achieve more accurate assessment of teaching quality. The results of simulation experiments show that the proposed method improves the F1 by 2.94–13.6%, which can accomplish a thorough and accurate evaluation of the effectiveness of teaching methods in database courses.

2 Relevant technologies

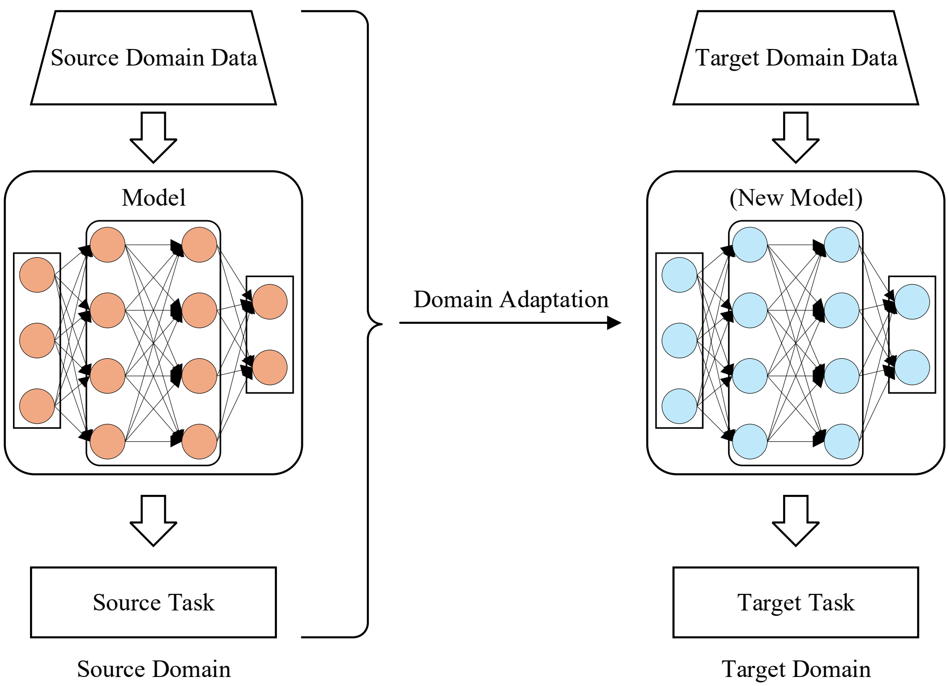
2.1 Domain adaptive

Domain adaptation (DA) belongs to an approach of TL, the issue under consideration is how to effectively migrate knowledge from one task or domain to a related one, with the goal of enhancing the model's overall performance (Zhou et al., 2021). In DA, two important concepts, i.e., source and objective fields, are often adopted to describe the different data and or domains involved in the training and application phases of the model, respectively. The source domain usually refers to a relatively rich and easily accessible dataset on which a model is trained to acquire certain knowledge, and the task of the origin field is to supply the model with adequate details so that it can adjust to the new objective domain in upcoming applications. The objective field is usually unseen outside the source field, so the target domain's data distribution might exhibit variations compared to the source domain's, and the objective of domain adaptive techniques is to leverage the knowledge acquired from the origin field to minimise the generalisation error in the objective domain.

DA refers to the capability of transferring a model from a origin field to an objective field through studying inter-domain discrepancies. The fundamental concept of DA is to construct a shared deep neural network model that operates across both the source and objective domains, and to facilitate the migration process from the source domain to the objective domain using feature transformation and model fine-tuning techniques (Singhal et al., 2023). However, in real life, there is not only one available origin field, and multi-source field DA aims to solve this problem by simultaneously adapting many origin fields to the objective field, as shown in Figure 1.

In multi-source DA, there are N various origin fields with distributions specified as $\{p_{s_j}(x, y)\}_{j=1}^{N}$, and marked origin field data $P_s = \{(X_{s_j}, Y_{s_j})\}_{j=1}^{N}$ are sampled from these distributions, individually, where $X_{s_j} = \{x_i^{s_j}\}_{i=1}^{|X_{s_j}|}$ denotes the sample of source domain j, and $Y_{s_j} = \{y_i^{s_j}\}_{i=1}^{|X_{s_j}|}$ is the corresponding true value label.

Figure 1 Schematic diagram of domain adaptation



In addition, there is an objective distribution $p_t(x, y)$ from which the objective domain data $X_t = \{x_i^t\}_{i=1}^{|X_t|}$ is sampled, but there is no corresponding real label Y_t. The target function of a DA is usually designed to minimise the divergence between the source and objective fields without compromising on the source domain's performance, as shown below:

$$L(\theta) = L_{src}(\theta) + \lambda D(\theta) \tag{1}$$

where $L(\cdot)$is the objective function, $L_{src}(\cdot)$is the loss on the source domain data, $D(\cdot)$ is an operation that measures the divergence in data between the source and target fields, and θ is a moderator term that balances the origin field loss with the field difference.

2.2 Transfer learning

TL is a vital technique in machine learning which aims to learn shared feature information from multiple relevant scenarios to improve the performance of AI algorithms in specific application scenarios (Lu et al., 2015). According to the different dimensions of 'how to migrate', TL can be categorised according to sample, feature, and metric learning. The sample-based approach extracts the most informative samples in the origin and objective fields, and reweights or resamples the sample set. The characteristic-based method is to detect the common attributes that persist across the source and objective domains, unaffected by domain shifts, learn the same feature representation, and reduce the distribution difference between the domains. The method based on metric studying is to shorten the distance among source and objective fields by adopting metrics such as maximum mean deviation, in light of the hypothesis that the origin and objective fields can be projected to a high-dimensional plane between the origin and objective fields.

It mainly involves two domains, the origin field, which is the field with a sufficient number of labelled data, and the objective field, which is the field with only sparse unmarked data. The objective field is usually a domain that has some similarity to the origin field and represents a different but linked process. The domain is the subject of transfer learning, and a domain comprises data x and a probability distribution $P(x)$ that defines the probabilities associated with these data points. The instance of data on the field includes inputs x and outputs y with a probability distribution denoted as $P(x, y)$. The probability distribution is the same as that on the domain. For any sample (x_i, y_i), there are $x_i \in X$, $y_i \in Y$, in which X stands for the characteristic and Y stands for the mark space. Origin field D_s and objective field D_t have various data distributions, i.e., $X_s \neq X_t$, $P(X_s) \neq P(X_t)$. The purpose of transfer learning is to adopt the origin field data to study a forecasting operation $f : x_t \mapsto y_t$ on the objective field, given D_s and D_t, such that f is with the smallest forecasting error E on the objective field.

$$f = \arg\min E_{(x,y)\in D_t}(f(x), y) \tag{2}$$

3 Text-embedded learning for BERT-based instructional feedback in database courses

The data for evaluating the teaching quality of database courses come from multiple sources of text data, including the teaching system, online platform learning data, and so on. This paper begins with an embedded study of feedback texts on the effectiveness of teaching and learning processes in a database course. The database course feedback texts were used to assess different aspects related to the quality of teaching, including teacher ethics, content, teaching attitude, teacher competence and learning environment. Sentence embedding learning based on BERT (Aum and Choe, 2021) is used for the feedback text and combined with topic word list embedding to lay the foundation for the development of the upcoming teaching quality assessment model.

Given an instructional feedback sentence S, $S = \{w_i \mid i \in [1, 2, \ldots, n]\}$, which contains n lexical elements w_i, each lexical element represents a basic linguistic unit in the sentence, which can be a complete Chinese character, a word, or a smaller subword, such as a single letter or a punctuation mark. The proposed approach aims at extracting the set A of terms for the pedagogical aspects of the mentioned m database courses, where $A = \{a_i | i \in [1, 2, \ldots, m]\}$. For the input sentence S, a unique categorisation label [CLS] appended to the start of the input sequence, followed by a separator [SEP] is appended to the end of the input sequence. Thus, the input sentence is altered to the sequence '[CLS]' at the beginning, followed by S, and ending with '[SEP]'.

The proposed method utilises the BERT language model to obtain word embeddings. Compared to language models such as Glove, BERT represents a profound, bidirectional model that is constructed upon the transformer architecture that generates dynamic context-dependent vectors. Moreover, BERT supports end-to-end fine-tuning, and the output layer can be adapted to text categorisation tasks by simply adjusting the output layer after pre-training. The words that have been preprocessed by the techniques of word segmentation, stemming, and deletion of deactivated words are fed into the BERT model to obtain a real-valued vector $b_w \in R^{d_w \times |\Omega|}$, where b_w is the matrix over the real number field (R) with dimension $d_w \times |\Omega|$, d_w is the embedding dimension, denoting the total columns within the matrix, and $|\Omega|$ is the lexicon size, denoting the number of rows in the matrix.

Given a context $C = \{c_i\}_{i=1}^{j} m$, j is the length of the context sentence. Based on the set of topic words in the context, construct an auxiliary sentence F for C. Thus, the input sequence S^* containing C and F is obtained.

$$S^* = [\langle CLS \rangle, C, \langle SEP \rangle, F, \langle SEP \rangle] \tag{3}$$

The embedding vector for a given lexical element w_i is denoted as follows:

$$h_i^0 = s_i^{tok} + s_i^{pos} + s_i^{seg} \tag{4}$$

where s_i^{tok}, s_i^{seg} and s_i^{pos} are the lexical element embedding vectors, interval embedding vectors and position embedding vectors of w_i, respectively. In the BERT model, s_i^{tok} represents the semantic information of each lexical element, and each word is mapped as a vector of fixed dimensions to help the model capture the semantic information. s_i^{seg} is used to distinguish between sentences or paragraphs, providing each word with information that identifies the sentence to which it belongs, and helping the model to distinguish the context of different sentences or paragraphs when processing long texts. s_i^{pos} indicates the position of each lexical element in the sentence, which is used to supplement the word order information to help the model understand the sentence structure.

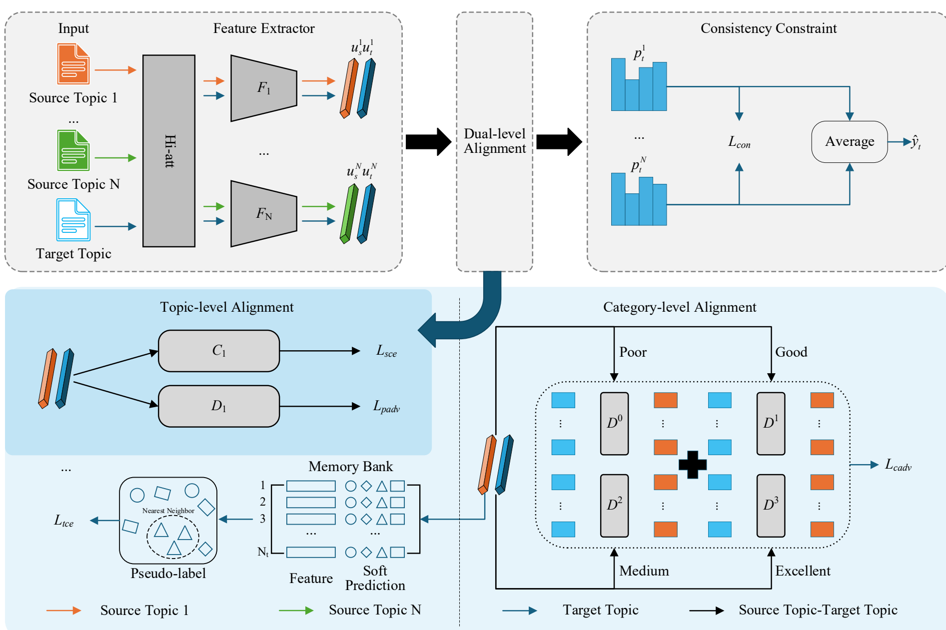
4 Evaluation of teaching quality in database courses based on domain-adaptive transfer learning

4.1 Design of teaching quality assessment model for database courses

After embedding the text of database course teaching quality assessment and obtaining the subject words of the text from multiple sources, this paper proposes a cross-topic database curriculum teaching quality automatic scoring approach in light of domain-adaptive transfer learning (DATL) to address the problem of the difficulty of distributional alignment among multiple source topics and the differences in the distribution of teaching quality scoring categories among different topics. As shown in Figure 2, DATL employs a two-layer adversarial network – topic adversarial migration network and category adversarial migration network – to achieve feature distribution migration alignment with regard to each combination of a source topic and a target topic. Moreover, DATL minimises the differences in ratings between classifiers through consistency constraints, thus providing a more accurate assessment of teaching quality.

DATL learns and aligns feature distributions synchronously at two levels: the topic level and the category level. Specifically, our method first targets each pair of origin and target topics and maps them to their respective feature spaces independently. Next, at the topic level, this paper realises macro-level alignment with the help of topic adversarial networks; at the category level, this paper realises category alignment at a finer level through category adversarial networks. The collaborative two-level alignment procedure enables the model to achieve the best possible pairing among the source and objective topics. With the goal of encouraging consistency across the board for each categoriser, we establish a uniformity requirement that works to minimise the output discrepancies between each categoriser pairs. Ultimately, we simultaneously optimise both topic-adversarial and category-adversarial networks during the process of learning shared characteristic depictions and categorisers.

Figure 2 Designed classification model for teaching evaluation of database courses (see online version for colours)



In this paper, we define the task of evaluating the teaching quality of a database course as a categorisation task, assuming that there is a collection of rated texts from N source topics, denoted as $S = \{P_1, P_2, \ldots, P_N\}$, where $P_i = \{(x_1^i, y_1^i), (x_2^i, y_2^i), \ldots, (x_m^i, y_m^i)\}$, P is the source topic *i*, (x_j^i, y_j^i) is the j^{th} evaluated text under the topic, and *m* is the number of texts under the topic. At the same time, there are a number of ungraded sets of text on the target topic that are labelled as $T\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, in which (x_t, y_t) is the t^{th} text under the target topic and *n* is the number of texts under the topic. The purpose of this research is to train a transferable model grounded in scored and unscored objective instructional evaluation texts. The goal of this paper is to learn a transferable model based on scored and unscored target instructional evaluation texts, in order to forecast a dependable score $\hat{y}_t = \phi(S, T; \theta)$ for the objective subject text, where θ is the model parameter which is to be determined through studying.

4.2 CNN-LSTM-based feedback text feature extraction

The characteristic extractor is designed to convert the source and target topic pairs into differing characteristic spaces, one for each. In this paper, CNN-LSTM is used to learn the representations of compositions and transform them to various characteristic spaces using a fully connected layer. For each sentence in the instructional feedback text, a CNN is used to transform its representation into an encoded form. The k^{th} text is characterised as follows:

$$z_k = f\left(W_n * [w_k : w_{k+n-1}] + b_n\right) \tag{5}$$

where w_k is the word vector representation, *f* is the activation operation, *n* corresponds to the size of the sliding window for the CNN architecture, W_n is the weight matrix, and b_n is the bias. Given that various words have varying impacts on teaching quality scores, by utilising word attention pooling, this paper aims to boost the effectiveness of sentence representation *s*, whose expressions are shown in equations (6)–(8).

$$\tilde{z}_k = \tanh\left(W_a * z_k + b_a\right) \tag{6}$$

$$a_k = \frac{\exp\left(W_z * \tilde{z}_k\right)}{\sum_{p=1}^{m} \exp\left(W_z * \tilde{z}_p\right)} \tag{7}$$

$$s = \sum_{k=1}^{m} a_k * z_k \tag{8}$$

where W_a and W_z are the weight matrices, b_a is the bias, $\tilde{z}_k$ is the attention vector for the k^{th} word and a_k is the attention weight for the k^{th} word, and *s* is the final sentence representation, which is the weighted sum of all words.

Teaching quality feedback text characterisation once the sentence characterisation has been derived, the LSTM model is utilised to gather contextual message. In order to obtain a global sequence of feature states for a text containing one sentences, all sentences are fed into the LSTM cell to obtain the final textual characterisation for teaching quality evaluation of a database course, as shown below:

$$h_q = LSTM\left(s_q, h_{q-1}\right) \tag{9}$$

where s_q and h_q are the input sentences and their hidden states in step *q*, individually.

Following this, sentence attention pooling is executed to boost text e's representation, as outlined here.

$$\tilde{h}_q = \tanh\left(W_{ha} * h_q + b_{ha}\right) \tag{10}$$

$$ha_q = \frac{\exp\left(W_h * \tilde{h}_q\right)}{\sum_{p=1}^{l} \exp\left(W_h * \tilde{h}_p\right)} \tag{11}$$

$$e = \sum_{q=1}^{l} ha_q * h_q \tag{12}$$

where W_{ha} and W_h are the weight matrices and b_{ha} is the bias term, $\tilde{h}_q$ is the attention vector for the q^{th} word, ha_q is the attention weight for the q^{th} word.

In particular, given the j^{th} textual feature h_s^j of the origin topic and the textual feature h_t from the target topic, the feature extractor F_j transforms them to a topic-specific characteristic space as follows, where F_j is a fully connected layer.

$$u_s^j = F_j\left(h_s^j\right), \; u_t^j = F_j\left(h_t\right) \tag{13}$$

4.3 Feedback text topic alignment based on domain adaptive transfer learning

In this paper, we obtain the teaching quality feedback texts in different feature spaces, in order to learn the shared features of each source and target topic text pair, we align their distributions at topic level and category level by adversarial network (Ding et al., 2023), this paper demonstrates the process of topic level alignment as follows.

Given a representation u_s^j of a source composition, it is fed into its corresponding topic-specific classifier C_j and discriminator D_j, respectively. C_j is a softmax classifier used to predict the label of each text. For the text of the j^{th} source topic, this paper computes the cross-entropy loss L_{ce} among their forecasted and actual scores.

$$L_{ce} = -\frac{1}{\left|S_j\right|} \sum_{s=1}^{\left|S_j\right|} \sum_{k=0}^{C-1} y_{s,k}^{j} \log C_{j,k}\left(u_s^j\right) \tag{14}$$

where $C_{j,k}(u_s^j)$ is the predicted probability of the evaluation text x_s^j in category k. The entire origin topic L_{ce} for N source topic-specific classifiers is therefore as follows. Therefore, for N origin topic-specific categorisers, the whole origin topic L_{ce} is as follows:

$$L_{sce} = \sum_{j=1}^{N} L_{ce} \tag{15}$$

The discriminator is also a softmax classifier to judge whether the input essays belong to the designated theme. For each pair of compositions on the origin and objective topics, this paper computes the related cross-entropy loss among their forecasted topic marks and real topic marks.

$$L_{pd} = -\frac{1}{|S_j|}\sum_{s=1}^{|S_j|}\log\left(1 - D_j\left(u_s^j\right)\right) - \frac{1}{|T|}\sum_{t=1}^{|T|}\log D_j\left(u_t^j\right) \tag{16}$$

When learning shared feature representations, topic-specific feature extractors work to minimise the discrepancy between predicted and actual categories of the source topic text for precise grading, while maximising the discriminator loss to confuse the discriminators. Therefore, in this paper, we call the total loss of N topic-specific discriminators to be the topic-layer adversarial loss, which is computed as equation (17).

$$L_{padv} = \sum_{j=1}^{N} L_{pd} \tag{17}$$

4.4 Alignment of instructional evaluation categories based on domain adaptive transfer learning

There were significant differences in the distribution of categories among the texts of the different themes. In order to take into account the inherent category structure of the different themes, texts within each pair of source and target themes were aligned at the category level. Before performing category-level alignment, we first generate pseudo-labels for unlabeled target topic compositions. In this paper, we craft a simple yet effective approach for generating pseudo-labels that encapsulates the inherent data structures at the local level by amalgamating neighbourhood details stored in a stored bank. The use of mnemonics allowed us to explore information-rich neighbours across the entire target set of topic texts, in contrast to the limitations observed in previous studies on small batches of data. Specifically, for each target evaluation text x_t in a small batch, compute its topic-specific feature representation u_t and soft prediction $p_t = C(u_t)$. To alleviate the problem of prediction ambiguity, the soft predictions are sharpened at a specific parameter $\tau = 1/2$ as follows, where $p_{t,k}$ is the prediction probability score of the k^{th} class.

$$\tilde{p}_{t,k} = \frac{p_{t,k}^{\tau}}{\sum_{t=1}^{|T|} p_{t,k}^{\tau}} \tag{18}$$

For the goal of storing the characteristic representations and forecasts of each objective topic text, this paper assigns to each topic a specific memory bank M. It is iteratively updated in a moving average manner, defined as follows, where λ is an update smoothing parameter.

$$\begin{cases} \tilde{u}_t = \lambda u_t + (1-\lambda)\tilde{u}_t \\ \tilde{p}_t = \lambda p_t + (1-\lambda)\tilde{p}_t \end{cases} \tag{19}$$

To determine the m nearest neighbours of x_t, this paper computes the cosine similarity among its characteristic representation and that of each text saved in M. To achieve the objective of combining the information from its m neighbouring entities, the soft labels of these neighbours are averaged to derive the soft label for entity x_t as follows:

$$\begin{cases} \hat{p}_t = \dfrac{1}{m} \sum\limits_{h \in N_t} \tilde{p}_t^h \\ \tilde{y}_t = \arg\max\limits_k \hat{p}_{t,k} \end{cases} \tag{20}$$

where N_t is the neighbourhood set of x_t. The maximum probability associated with it is assigned as the confidence level w_t for the pseudo-label $\tilde{y}_t$. Next, the loss of all target topic texts is computed by the L_{ce} of the weighted confidence in light of N topic-specific categorisers.

$$L_{tce} = -\frac{1}{T} \sum_{t=1}^{|T|} \sum_{j=1}^{N} \sum_{k=0}^{C-1} \omega * \tilde{y}_{t,k}^j \log C_j \left(u_t^j\right) \tag{21}$$

With the aim of pairing up every source topic with its respective target topic, we evaluate how textual category-level characteristics are distributed, texts categorised under the same class are input into their respective class discriminators. The category discriminators are designed to judge whether the provided compositions pertain to the objective topic. Therefore, in this paper, the L_{ce} of the k^{th} discriminator is calculated in the following way:

$$L_{cd,k} = -\frac{1}{\left|S_{j,k}\right|} \sum_{u_s \in S_{j,k}} \log\left(1 - D_{j,k}\left(u_s^j\right)\right) - \frac{1}{|T_k|} \sum_{u_t \in T_k} \log D_{j,k}\left(u_t^j\right) \tag{22}$$

where $S_{j,k} \subset S_j$ and $T_k \subset T$ represent the set of database course instructional evaluation texts for the j^{th} source topic and the k^{th} category in the target topic, respectively.

It is worth noting that in order to align the category-level feature distributions of the instructional feedback text for all pairs of origin and objective topics, the topic-specific feature extractor attempts to obfuscate its corresponding category-level discriminator. Therefore, in this paper, the overall loss of the class-level identifier for N origin and objective topic pairs is referred to as the class-level hostile loss using the following equation:

$$L_{cadv} = \sum_{j=1}^{N} \sum_{k=0}^{C-1} L_{cd,k} \tag{23}$$

4.5 Classifier consistency constraint

After completing the topic-level and class-level alignment for all pairs of origin and objective topics, N predictions will be obtained for each target topic text. Drawing from the literature (Xu et al., 2022), this paper introduces consistency constraints to encourage consistency among these N topic-specific classifiers. Equation (24) is used to compute the whole value of the difference among the prediction probabilities generated through all pairs of topic-specific categorisers.

$$L_{con} = \frac{2}{N \times (N-1)} \sum_{t=1}^{|T|} \sum_{j=1}^{N-1} \sum_{p=j+1}^{N} \left| C_j\left(u_t^j\right) - C_p\left(u_t^j\right) \right| \tag{24}$$

In summary, the final loss in this paper consists of the classification loss, the double adversarial loss, and the categoriser uniformity loss. Considering that DATL obtains shared characteristic representations by the process of hostile training, this implies that during parameter updates, the discriminator and categoriser move in opposite gradient directions. To cope with the above issue, this study implements the backpropagation progress through mechanically inverting the gradient direction associated with the discriminator loss prior to its propagation to topic-specific feature extractor parameters, thereby gaining hostile training effects. Thus, the ultimate loss equals the aggregation of these individual losses, as shown below, where α, β are the weighting factors that govern the comparative influence of distinct losses.

$$L = L_{sce} + L_{tce} + \alpha\left(L_{padv} + L_{cadv}\right) + \beta L_{con} \tag{25}$$

Upon thoroughly refining these loss functions, an optimal dual adversarial migration network is achieved. For each text assessing teaching quality, the concluding score can be obtained through averaging the forecasts of each origin categorisers tailored to specific topics.

5 Experimental results and analyses

In this paper, the text data of teaching quality evaluation of a university database course is used as the experimental dataset, and the training dataset is constructed by manual annotation, and each data sample consists of students' or experts' evaluation contents and corresponding evaluation scores. After cleaning and removing dirty data, a total of 4,125 evaluation text data were obtained. The experimental hardware platform uses Intel Core i5-12400 CPU, NVIDIA GeForce RTX 3060 GPU, and 32 GB of RAM. Python 3.7.13 was chosen as the programming language, and CUDA version 11.3 was used to build the models, based on the compatible PyTorch version 1.11.0 deep studying structure.

In the experiments, this paper uses 50-dimensional BERT word embeddings, and in the configuration of a CNN designed for generalised characteristic extraction, this paper employs 100 filters. The obscured level dimension of the LSTM, which was initially obscured, was likewise configured to 100. During training, we configured the batch size to 32, the initial learning rate to 0.001, and the maximum number of training epochs to 600. Additionally, the dropout rate was adjusted to 0.5.

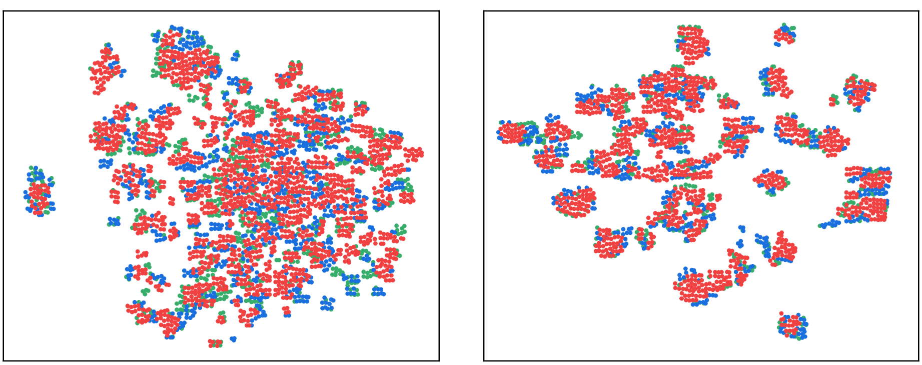
In this paper, the real ratings of five database course instructors (T1, T2, T3, T4, T5) are selected as a control and the teaching quality ratings are predicted using the models proposed in this paper, DATL, BERT-GRU (Sebbaq and El Faddouli, 2022), CNN-LSTM (Mao et al., 2024), CNN-TL (Yun et al., 2023) and WKETL (Guo, et al., 2024), and the results are shown in Table 1. The predicted values of DATL for teaching quality scores are highly consistent with the real values, with an error of 0.61 points or something like that, the predicted values of BERT-GRU deviate more from the real values, with a forecasting error of at least 4.78 points, and the predicted values of CNN-LSTM are also in some error with the true values, with a forecasting error of at least 4.54 points. In comparing CNN-TL and WKETL, the predicted value of CNN-TL is obviously deviating from the real value, with a forecasting error of at least 2.41, and the predicted value of WKETL is closer to the real value, with a forecasting error of less than

2.5 points, and the above analysis suggests that the designed DATL has a better impact of evaluating teaching quality.

Table 1 Predicted outcome for teaching quality ratings in database courses

Teacher	*Actual score*	*Predicted score*				
		BERT-GRU	*CNN-LSTM*	*CNN-TL*	*WKETL*	*DATL*
T1	73	65.39	78.15	70.59	74.18	73.21
	88	93.01	92.54	84.36	90.51	87.95
T2	89	85.22	86.35	92.51	88.07	89.12
	66	71.54	61.05	69.77	68.94	66.05
T3	75	82.54	70.06	71.23	72.01	74.84
	84	89.04	79.16	88.64	85.38	83.29
T4	93	86.18	88.03	96.27	95.21	93.61
	81	87.92	85.09	77.34	78.96	80.57
T5	72	78.05	77.93	75.94	70.19	72.37
	96	90.81	91.52	92.05	98.56	96.08

Figure 3 Categorical comparison of instructional evaluation results before and after DA, (a) before domain adaptive migration learning (b) after domain adaptive migration learning (see online version for colours)



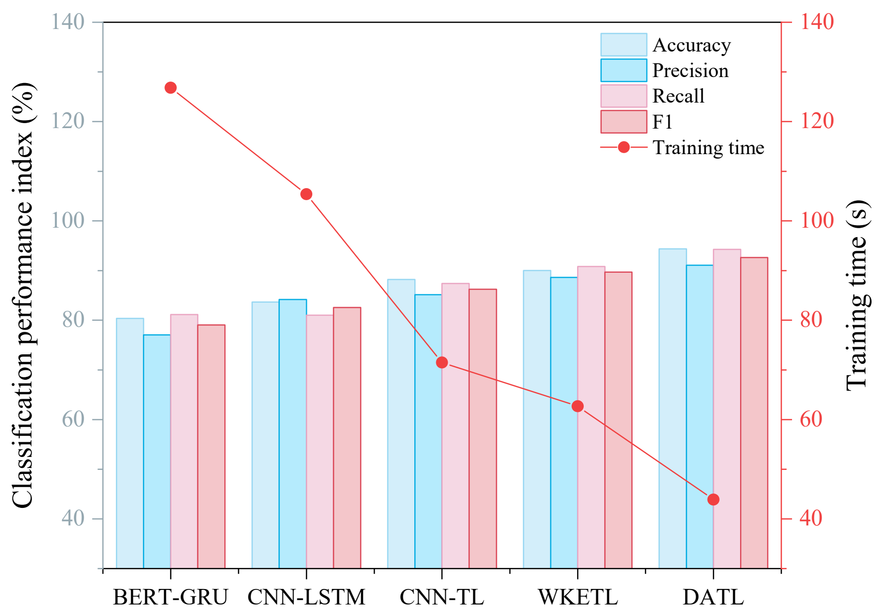
(a) (b)

The visualised comparison of the classification of the proposed method in this paper before and after domain adaptation is shown in Figure 3, where the results of various types of teaching quality evaluations are more dispersed before DA, and it is difficult to identify which category the final teaching evaluation results are. After DA, the results of each type of teaching evaluation of DATL are clearly visible, indicating that DATL has better classification performance. DATL not only designs the DA adversarial migration network to achieve the alignment of characteristic distributions of all pairs of source and objective topics, but also minimises the differences in ratings between different classifiers through consistency constraints, so as to more accurately assess the quality of teaching.

Comparison of classification accuracy, precision, recall, F1 and training time of different methods are shown in Figure 4. DATL has a classification accuracy of 94.39%, which is improved by 14.02%, 10.7%, 6.18% and 4.37% compared to BERT-GRU,

CNN-LSTM, CNN-TL and WKETL, respectively. Comparing the combined index F1 of precision and recall, DATL's F1 is 92.65%, which is improved by 2.94–13.6% compared with the other four methods, indicating that DATL is more efficient in categorising the results of the evaluation of teaching quality. Comparing the training time again, DATL has the shortest training time, which is 82.9 s, 61.5 s, 27.6 s and 18.8 s shorter than BERT-GRU,
CNN-LSTM, CNN-TL, and WKETL, respectively. Although ERT-GRU and CNN-LSTM realise the classification of teaching quality evaluation results through deep learning models, they do not consider the migration learning problem of multi-source evaluation text data, so the classification effect is not good. CNN-TL combines the ideas of CNN and migration learning to classify the results of teaching quality evaluation, but it does not consider the DA problem of multi-source data, so the classification accuracy is low. Although WKETL considers the DA problem for multi-source evaluation data, it does not align the subject terms and evaluation categories in the teaching evaluation text, which results in a less accurate classification than DATL. In summary, DATL achieves an accurate assessment of the teaching quality of database courses.

Figure 4 Comparison of classification performance metrics and training time (see online version for colours)



6 Conclusions

As an important basic course in computer science, the assessment of the teaching quality of database courses is vital to improve the teaching effect. However, there are variations in the distribution of instructional data, and traditional methods often struggle to deal effectively with such domain variations, leading to poor classification accuracy in evaluation models. To this end, this paper first utilises BERT model for embedded learning of teaching feedback text, which is used to assess different aspects related to teaching quality. Then the local features of the feedback text are extracted by CNN and the global features of the feedback text are extracted by LSTM, and the text features are enhanced by the attention mechanism. Based on this, for the goal of studying the shared features of each source and objective topic text pair, this paper generates pseudo-labels for untagged target topic texts by aligning their distributions at the topic level and category level through domain adaptive transfer learning. The data structure inherent within a local level is captured by aggregating neighbourhood information to minimise the differences in scores across classifiers and achieve more accurate assessment of teaching quality. The simulation outcome demonstrate that the designed approach acquires significant performance enhancement on the task of teaching quality assessment of database courses compared with the traditional method, which provides a new idea to improve the accuracy and reliability of teaching quality evaluation.

Although the results of this article have achieved a good accuracy rate of teaching quality evaluation, there always exists more space to be explored and improved in the journey of scientific inquiry:

1. The model's ability to generalise is a focus for future research. Despite good results on existing test sets, this paper needs to further explore the model's adaptability to more unknown topics and datasets to ensure that the accuracy and reliability of its scores are not limited to a specific sample set.
2. Interpretability of the model is also an important aspect that will be studied in depth in the future of this paper. As a black-box system, the model in this paper needs more transparency to ensure that users have a clear understanding of the model's judgement and decision-making process, which is important for the trust and acceptance of the model.

Overall, although the research in this paper has been accomplished in some aspects, this paper is conscious of the fact that there are still many important issues to be deeply investigated. This paper will continue to optimise the methodology of this paper and expand the boundaries of the study, with the aim of responding more fruitfully to the challenges posed in the future, and advancing the knowledge and technology of evaluating the quality of teaching and learning in database courses moving forward.

Declarations

All authors declare that they have no conflicts of interest.

References

Aum, S. and Choe, S. (2021) 'srBERT: automatic article classification model for systematic review using BERT', *Systematic Reviews*, Vol. 10, pp.1–8.

Chen, L. and Mokhtar, M.M. (2023) 'Education on quality assurance and assessment in teaching quality of high school instructors', *Journal of Big Data*, Vol. 10, No. 1, p.142.

Ding, H., Sun, Y., Huang, N., Shen, Z., Wang, Z., Iftekhar, A. and Cui, X. (2023) 'RVGAN-TL: a generative adversarial networks and transfer learning-based hybrid approach for imbalanced data classification', *Information Sciences*, Vol. 629, pp.184–203.

Dong, Y. (2023) 'Teaching quality monitoring and evaluation in higher education through a big data analysis', *International Journal of Emerging Technologies in Learning (Online)*, Vol. 18, No. 8, p.61.

Guo, S., Yang, X., Farizan, N.H. and Samsudin, S. (2024) 'The analysis of teaching quality evaluation for the college sports dance by convolutional neural network model and deep learning', *Heliyon*, Vol. 10, No. 16,.

Hou, J. (2021) 'Online teaching quality evaluation model based on support vector machine and decision tree', *Journal of Intelligent & Fuzzy Systems*, Vol. 40, No. 2, pp.2193–2203.

Ishaq, M., Abid, A., Farooq, M.S., Manzoor, M.F., Farooq, U., Abid, K. and Helou, M.A. (2023) 'Advances in database systems education: methods, tools, curricula, and way forward', *Education and Information Technologies*, Vol. 28, No. 3, pp.2681–2725.

Lagus, J., Longi, K., Klami, A. and Hellas, A. (2018) 'Transfer-learning methods in programming course outcome prediction', *ACM Transactions on Computing Education (TOCE)*, Vol. 18, No. 4, pp.1–18.

Li, S. and Yuan, X. (2021) '[Retracted] Application of linear regression mathematical model in the evaluation of teachers' informatization quality', *Complexity*, Vol. 13, No. 1, pp.55–68.

Li, Y. (2022) 'Deep learning-based correlation analysis between the evaluation score of English teaching quality and the knowledge points', *Computational Intelligence and Neuroscience*, Vol. 20, No. 4, pp.41–54.

Li, Y. and Zhang, H. (2024) 'Big data technology for teaching quality monitoring and improvement in higher education-joint K-means clustering algorithm and apriori algorithm', *Systems and Soft Computing*, Vol. 6, pp.1–16.

Liu, Y. (2022) 'Improved convolutional neural networks for course teaching quality assessment', *Advances in Multimedia*, Vol. 18, No. 1, pp.13–29.

Lu, J., Behbood, V., Hao, P., Zuo, H., Xue, S. and Zhang, G. (2015) 'Transfer learning using computational intelligence: a survey', *Knowledge-Based Systems*, Vol. 80, pp.14–23.

Mao, Y., Wang, F., Zhang, Y. and Chen, Q. (2024) 'Teaching quality evaluation model based on CNN-framework with entropy weight method', *Computer-Aided Design and Applications*, Vol. 22, pp.297–312.

Musti, K.S. (2015) 'An effective approach for teaching database course', *International Journal of Learning, Teaching and Educational Research*, Vol. 12, No. 1, pp.53–63.

Sebbaq, H. and El Faddouli, N-e. (2022) 'An explainable attention-based bidirectional GRU model for pedagogical classification of MOOCs', *Interactive Technology and Smart Education*, Vol. 19, No. 4, pp.396–421.

Singhal, P., Walambe, R., Ramanna, S. and Kotecha, K. (2023) 'Domain adaptation: challenges, methods, datasets, and applications', *IEEE Access*, Vol. 11, pp.6973–7020.

Tsiakmaki, M., Kostopoulos, G., Kotsiantis, S. and Ragos, O. (2020) 'Transfer learning from deep neural networks for predicting student performance', *Applied Sciences*, Vol. 10, No. 6, p.2145.

Wang, H., Yang, F. and Xing, X. (2022a) 'Evaluation method of physical education teaching and training quality based on deep learning', *Computational Intelligence and Neuroscience*, Vol. 8, No. 3, pp.16–31.

Wang, Z., Xu, T. and Wang, L. (2022b) 'Teaching evaluation method based on fuzzy support vector machine algorithm', *Mobile Information Systems*, Vol. 16, No. 5, pp.73–85.

Xu, B., Yan, S., Li, S. and Du, Y. (2022) 'A federated transfer learning framework based on heterogeneous domain adaptation for students' grades classification', *Applied Sciences*, Vol. 12, No. 21, pp.10–17.

Yun, G., Ravi, R.V. and Jumani, A.K. (2023) 'Analysis of the teaching quality on deep learning-based innovative ideological political education platform', *Progress in Artificial Intelligence*, Vol. 12, No. 2, pp.175–186.

Zhou, K., Yang, Y., Qiao, Y. and Xiang, T. (2021) 'Domain adaptive ensemble learning', *IEEE Transactions on Image Processing*, Vol. 30, pp.8008–8018.