International Journal of Knowledge-Based Development

ISSN online: 2040-4476 - ISSN print: 2040-4468
https://www.inderscience.com/ijkbd

Research on a recommendation model for sustainable innovative teaching of Chinese as a foreign language based on the data mining algorithm

Yingying Zhang, Huiyu Guo

DOI: 10.1504/IJKBD.2023.10056110

Article History:

Received:	20 December 2022
Last revised:	03 February 2023
Accepted:	04 March 2023
Published online:	27 March 2024

Research on a recommendation model for sustainable innovative teaching of Chinese as a foreign language based on the data mining algorithm

Yingying Zhang* and Huiyu Guo

International Education College,
Gannan Medical University,
Ganzhou, 341000, China
Email: zhangyy.zhang@yandex.com
Email: dongdu1979@163.com
*Corresponding author

Abstract: With the continuous development of teaching Chinese as a foreign language, more teaching methods are combined with network teaching. However, it is difficult for network teaching methods to find ways that are suitable for different learners from various teaching resources. Therefore, to help learners obtain appropriate teaching methods from the network teaching platform, the research establishes a network teaching recommendation model for Chinese as a foreign language based on the user's interest similarity. Three experimental schemes are designed to verify the effect of the proposed model. The experimental results show that the mean absolute error (MAE) scores of the model in the three schemes are 0.67, 0.7095, and 0.7428, respectively; the RMSE scores are 0.88, 0.9346, and 0.9695, respectively. Thus, the proposed collaborative filtering recommendation algorithm based on user interest similarity migration has good recommendation performance.

Keywords: data mining; transfer learning; Chinese as a foreign language; teaching innovation; collaborative filtering.

Reference to this paper should be made as follows: Zhang, Y. and Guo, H. (2024) 'Research on a recommendation model for sustainable innovative teaching of Chinese as a foreign language based on the data mining algorithm', *Int. J. Knowledge-Based Development*, Vol. 14, No. 1, pp.1–18.

Biographical notes: Yingying Zhang obtained her Master degree in Chinese linguistic (2012) from Jiangxi normal university, China. Presently, she is working as a Lecturer and the Vice Dean in the Chinese Department of International Education College, Gannan Medical University, China. She was invited as a resource person to deliver various teaching talks on teaching Chinese as the second language for foreign students. She is also serving and served as a reviewer for national, international conferences and journals. She has published 5 articles in Chinese journals and conferences proceedings. Her areas of interest include teaching Chinese as the second language, international education, communication and cooperation, advanced education.

Huiyu Guo, Lecturer; From September 2005 to July 2009, she graduated from Huanghuai University, majoring in Business English and obtained a Bachelor's degree; From September 2009 to July 2012, she graduated from Shanghai Normal University with a Master's degree in English language and literature; Now, she works in the International Education College, Gannan Medical University, Lecturer; The academic research direction is studying abroad in

China; She has published seven academic articles and presided over three municipal projects.

1 Introduction

As China's national strength continues to increase, the international exchange of Chinese culture has become more frequent, and the number of friends learning Chinese has also increased. Due to the limitations of traditional Chinese teaching in regions and resources, it is difficult to meet the rapid development of Chinese as a foreign language (CFL). Therefore, the teaching of Chinese as a foreign language (TCFL) needs to maintain continuous and innovative teaching (Nguyen, 2021; Saura et al., 2023; Ribeiro-Navarrete et al., 2021). Network teaching has obvious advantages that eliminates the geographical and resource factors of traditional teaching for teaching CFL. Many students have gradually adapted to the online TCFL. While online teaching has brought convenience for learners, it has also created some new problems. There are a lot of teaching resources in online teaching, but it is difficult for different learners to confirm the teaching content or teaching method that suitable for themselves, which has a serious impact on the teaching effect (Zhang and lynch, 2021; Barbosa et al., 2022; Saura et al., 2021). Network teaching is often lack of interaction, the classroom is relatively boring, and the teaching content cannot be flexibly adjusted according to the learning quality of learners. How to solve these problems is a key to promote the development of Chinese language teaching, and also the main goal of the research. Therefore, based on different learners' behaviors and preferences, it is important to recommend suitable online teaching methods for them (Zeng and Jiang, 2021). The research proposes a user similarity transfer for collaborative filtering (UST-CF) algorithm based on the migration of user interest similarity. The core of this algorithm is to utilise the user interest similarity in other fields to help the target field learning the user interest similarity, and combine the collaborative filtering algorithm to achieve the classification recommendation of projects in the target field. Compared with the traditional collaborative filtering recommendation model, this model has the characteristics of self-regulation of parameter size, which makes the model recommendation effect more accurate. In the UST-CF model, the balance parameters affecting the model are analysed to obtain the approximate distance estimation method of feature subspace. By filling in the missing matrix in the auxiliary domain, the user interest similarity in the auxiliary domain is successfully obtained. Then the user interest similarity is migrated to the target domain, and finally the user interest similarity in the target domain is obtained, and the recommendation list is generated to complete the user interest recommendation. The research has alleviated the problem of data scarcity and cold start in the recommendation algorithm, so that the courses of CFL can be recommended according to the habits of different customers. In order to improve the quality and accuracy of foreign language recommendation, the main innovation of the UST-CF model is to use subspace distance to estimate the balance parameters of the model, which improves the intelligence of the model. The method to be used in the study is compared with other collaborative filtering methods, and the comparison results are shown in Table 1.

Table 1 Comparison of representative algorithms of collaborative filtering technology

Collaborative filtering strategy	*Representative algorithm*	*Advantage*	*Shortcoming*
Memory-based CF (Grl et al., 2020)	User based collaborative filtering Project based collaborative filtering	Simple application; It is unnecessary to consider the attributes of the recommended items; High degree of automation	Depends on user ratings; When the data is sparse, the accuracy decreases; Cannot recommend new users and sex items
Model-based CF (Iwendi et al., 2022)	Collaborative filtering based on bayesian trust network Collaborative Filtering Based on Latent Semantics Collaborative filtering based on clustering Collaborative filtering based on dimension reduction	Effectively deal with problems such as sparsity; Improve recommendation accuracy; Provides intuitive use of recommended effects	The establishment of the model costs more resources; Trade off between prediction accuracy and scalability; Dimension reduction technology is easy to lose useful information
Hybrid recommenders (Fan et al., 2021)	Content Based Collaborative Filtering Content based collaborative filtering Memory based and model-based collaborative filtering	No cold start problem; The recommended precision is high; Eliminate data sparsity and grey users	High implementation and application costs; It is difficult to obtain additional knowledge

2 Related work

In the continuous development of CFL, many scholars have conducted more in-depth research on the innovation of teaching methods and the recommendation of teaching intelligence to promote the teaching of CFL with higher quality. Yang et al. (2018) used flipped learning as a contrast in CFL. Through quantitative and qualitative data analysis, it is found that there are significant differences in learning outcomes and teaching satisfaction between traditional teaching classes and innovative teaching classes. The experimental results show that the innovation of teaching methods is conducive to improving learners' interest, thereby improving the quality of learning. Yu and Geng (2020) found that there was a high dropout rate at the Chinese character learning stage in CFL. Therefore, the research carried out innovation from the Chinese character teaching in CFL, analysed the importance of Chinese characters in Chinese, and fully solved the problem of teaching and learning in teaching. Gong et al. (2018) put forward suggestions for cooperation with mainland China to promote the continuous development of Chinese language education. Attaran and Yishuai (2018) analysed the impact of CFL by studying in-service teachers. The experimental results showed that the lack of sufficient teaching experience leads to the decline of the teaching quality of CFL. According to the shortcomings of the problem, the study gave targeted suggestions on building teachers' professionalism and identity to improve the efficiency and quality of curriculum

education. Li (2021) proposed the IRF classroom discourse structure model, applied it in the classroom teaching of CFL, and made a practical analysis of the Chinese corpus. It was concluded that the model was applicable to Chinese classroom teaching. The research results have effectively helped Chinese teachers to improve the quality of Korean teaching and put forward constructive suggestions for Chinese teaching. Li et al. (2019) innovated in Chinese vocabulary learning in CFL. By comparing electronic flashcards with paper flashcards, this paper analysed the students' learning situation of Chinese vocabulary and the change of learning attitude. Through relevant experiments, the results showed that after the innovative electronic flashcard learning, students have achieved better results in word reading and listening tests. Therefore, this innovative approach to CFL helps students showed a more positive learning attitude (Li and Tong, 2019).

Chen et al. (2018) proposed a disease diagnosis and treatment recommendation system, which can diagnose atypical symptoms of diseases. The system introduced the DPCA algorithm to improve the accuracy of disease symptom recognition and used a data mining algorithm to perform association analysis on disease diagnosis rules and treatment rules according to symptoms, to achieve disease symptom clustering and provide intelligent treatment opinions. Xing et al. (2022) conducted multi-dimensional data analysis on the information dissemination of COVID-19 in the network and provided a differentiated assessment of the evolution of the COVID-19 epidemic in the network through data mining and text clustering methods. The differentiated assessment clearly showed the characteristic changes in public opinion in different periods, successfully predicted people's topics of discussion in the network, and played a preventive role in the dissemination of emergency information. Schwalbe-koda et al. (2022) used a data mining algorithm as a guiding agent for zeolite synthesis of organic structures. The experiment of KFI zeolite as an example showed that the experiment of organic synthesis of this zeolite was successful under the data mining technology. Mosavi et al. (2022) used CRISP-DM method to mine the data collected by industrial sensors, predicted the power consumption of Bosch automotive multimedia facilities, and analysed the impact of temperature and humidity on power consumption. The experiment showed that data mining technology successfully predicted power consumption, and combined with industrial management system to achieve the effect of intelligent management. Envelope (2021) built an automatic identification model of leakage users based on data mining technology. Through analysis, the results proved that the model had a high recognition accuracy, and promoted the inspection of power facilities. Jin and Hu (2022) used data mining technology to conduct a digital transformation of energy enterprises when solving problems in rural energy investment and financing systems.

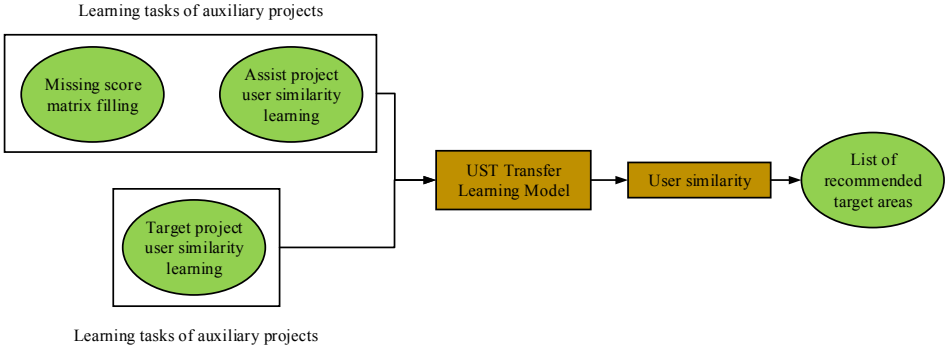
Based on the above analysis, data mining has a wide range of applications in the internet, automotive, and other fields, especially in the fields of prediction, classification, etc. As far as teaching CFL is concerned, many scholars have made innovations in teaching forms in different ways, but few have applied data mining technology to teaching methods. Therefore, the research has been made to use data mining techniques to build a personalised recommendation model for teaching CFL, to help learners get more suitable and efficient learning methods.

3 Construction of collaborative filtering recommendation algorithm based on user interest similarity migration

3.1 Learning task analysis of auxiliary items

The most commonly used method in collaborative filtering algorithms is to recommend the things that the nearest neighbour is interested in. This method is called the nearest neighbour method (Ajaegbu, 2021; Na et al., 2021). Therefore, the quality of object recommendation mainly depends on the selection of the nearest neighbour. It is based on the similarity between users. The strength of interest relevance between users is related to the similarity value, and the similarity calculation depends on the user's rating of the object item (Han et al., 2021). Therefore, the research proposes a user similarity transfer for collaborative filtering (UST-CF) algorithm based on user interest similarity migration. The flow of the UST-CF algorithm is shown in Figure 1.

Figure 1 Specific flow chart of UST-CF algorithm (see online version for colours)



In Figure 1, the learning task is divided into two parts by the UST-CF algorithm, one of which is the learning task of the auxiliary project, and the other is the learning task of the target project. Assuming that the common user set of the auxiliary project and the target project is U_C, the similarity of the auxiliary project users is defined as $A\_sim(u_i,u_j)$, and the similarity of the target project users is defined as $T\_sim(u_i,u_j)$; the final target project similarity $U\_sim(u_i,u_j)$ is obtained by weighting $A\_sim(u_i,u_j)$ and $T\_sim(u_i,u_j)$. The model expression is shown in formula (1).

$$U\_sim(u_i,u_j) = \alpha A\_sim(u_i,u_j) + (1-\alpha)T\_sim(u_i,u_j) \tag{1}$$

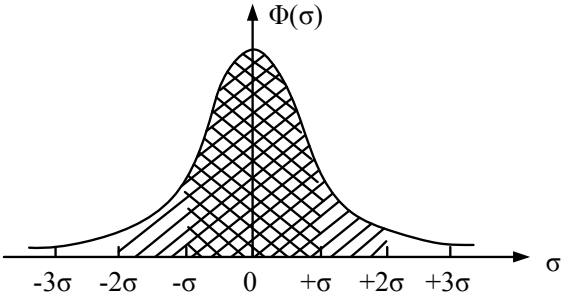
Ins formula (1), α represents the balance parameter, and the value range is [0,1), which is commonly used to adjust the migration degree of interest similarity. When the value of α is larger, the migration degree of the target project using auxiliary projects is more; when the value of α is smaller, the migration degree is less. When the value of α is 0, it means that the target project does not use the information of auxiliary items at all. In the recommendation system, the user-item scoring matrix often has missing values. The

average filling (AF) is the most basic method to fill in the missing matrix. The filling form is shown in formula (2).

$$\begin{cases} \hat{r}_{u.} = \overline{r}_{u.} \\ \hat{r}_{ui} = \overline{r}_{.i} \\ \hat{r}_{ui} = (\overline{r}_{u.} + \overline{r}_{.i}) / 2 \\ \hat{r}_{ui} = b_{u.} + \overline{r}_{.i} \\ \hat{r}_{ui} = \overline{r}_{u.} + b_{.i} \\ \hat{r}_{ui} = \overline{r} + b_{u.} + b_{.i} \end{cases} \tag{2}$$

In formula (2), $\overline{r}_{u.}$ represents the average user rating; $\overline{r}_{.i}$ represents the average score of the project; $b_{u.}$ and $b_{.i}$ represent user and project scoring preferences respectively; $\overline{r}$ represents the average value of the whole scoring matrix. AF method enhances the stability of similarity calculation, but weakens the data characteristics of the coefficient matrix to a certain extent. In the collaborative filtering algorithm of matrix decomposition, the algorithm can fill in the missing part of the scoring matrix by decomposing the matrix, so that the low-rank property of the coefficient matrix can be guaranteed. Therefore, the prediction accuracy of the model is strengthened. The matrix with noise data is represented by R_A, and the distribution mode of R_A is Gaussian distribution with equal direction. Its distribution form is shown in Figure 2.

Figure 2 Gaussian distribution map



In Figure 2, σ represents the dispersion degree of normal distribution data. The optimal solution of the characteristic matrix of the user and the project can be calculated through the loss function, and the loss function expression is shown in formula (3).

$$\Gamma(U,V) = \left\| R_A - UV \right\|_F^2 \to \min \tag{3}$$

In formula (3), U and V represent the characteristic matrix of users and projects respectively. To find the optimal solution for formula (3), it is actually to solve the low rank matrix and make the low rank matrix close to the original scoring matrix. In the singular value decomposition (SVD) of the matrix, the relationship between the original scoring matrix and the user orthogonal matrix as well as the item orthogonal matrix is shown in formula (4) (Li et al., 2021).

$$R_A = U \times S \times V^T \tag{4}$$

In formula (4), T represents matrix transposition; S represents a diagonal matrix. The expression of diagonal matrix is shown in formula (5).

$$\begin{cases} S = \begin{bmatrix} S_1 & 0 \\ 0 & 0 \end{bmatrix} \\ S_1 = diag(\sigma_1, \sigma_2, \cdots, \sigma_r) \end{cases} \tag{5}$$

In formula (5), the elements of matrix S_1 are arranged in the order of largest to smallest, and S represents the rank of the original scoring matrix. If the first d largest singular values of the original scoring matrix form a diagonal matrix, there is $S_d = diag(\sigma_1, \sigma_2, \ldots, \sigma_d)$. The orthogonal matrix $U_{p\times d}$ and $V_{q\times d}^T$ are composed of the left and right singular vectors corresponding to the elements in the diagonal matrix S_d. Then the filling matrix Z of auxiliary items is shown in formula (6).

$$Z = U_{p\times d} \times S_{d\times d} \times V_{q\times d}^T \tag{6}$$

After the filling matrix of auxiliary items is obtained, the user's prediction score for the item can be calculated. The scoring calculation method refers to formula (7).

$$p_{i,j} = \overline{r_i} + U_{i.} \times S_{d\times d} \times V_{j.}^T \tag{7}$$

In formula (7), $p_{i,j}$ represents the predicted score of item v_j in user u_i. Since R_A is a missing matrix, if missing factors are not considered, the loss function of equation (3) has obvious deviation in the final filled value. Therefore, the loss function should be modified to equation (8) without considering missing factors.

$$\Gamma(U,V) = \left\| W_A \odot (R_A - UV) \right\|_F^2 \to \min \tag{8}$$

In formula (8), W_A is the marking matrix. To prevent the loss function from overfitting, the regular terms of user characteristic matrix and project characteristic matrix are introduced into formula (8) to obtain the matrix decomposition model after reconstruction. The specific expression is shown in formula (9).

$$\Gamma(U,V) = \left\| W_A \odot (R_A - UV) \right\|_F^2 + \lambda_1 \left\| U \right\|_F^2 + \lambda_2 \left\| V \right\|_F^2 \to \min \tag{9}$$

In formula (9), λ is the control parameter of introducing the regularisation term to adjust the training error between the filled matrices. The iterative algorithm of Alternation Least Squares (ALS) is used to find the optimal solution of formula (9) (Du et al., 2019). Keep V unchanged and take the derivative of $U_{i.}$ to obtain the solution formula of $U_{i.}$, as shown in formula (10).

$$U_{i.} = R_{Ai.} V_{u.i} (V_{u.j}^T V_{u.i} + \lambda_1 n_{u.i} I)^{-1} \tag{10}$$

In formula (10), $R_{A.j}$ represents the scoring vector of common users who have evaluated the item $\left| v_j \right|$; $U_{v.j}$ represents the feature vector of the common users who have evaluated item $\left| v_j \right|$; $\left| n_{v,j} \right|$ represents the number of users who have evaluated project $\left| v_j \right|$, and I

represents the unit matrix. Keep U unchanged and take the derivative of V_j to get the solution formula of V_j, as shown in formula (11).

$$V_j = R_{A.j} U_{v.j} (U_{v.j}^T U_{v.j} + \lambda_2 n_{v.j} I)^{-1} \tag{11}$$

Formulas (10) and (11) update the row vectors of U and V respectively until the algorithm converges. Finally, the filled matrix can be obtained. In the auxiliary project, in addition to filling in the missing scoring matrix, it is also necessary to calculate the common user similarity. The specific calculation formula is shown in formula (12).

$$A\_sim(u_i, u_j) = \frac{\sum_{k \in I} (r_{i,k} - \overline{r}_i)(r_{j,k} - \overline{r}_j)}{\sqrt{\sum_{k \in I} (r_{i,k} - \overline{r}_i)^2} \sqrt{\sum_{k \in I} (r_{j,k} - \overline{r}_j)^2}} \tag{12}$$

3.2 Learning task analysis of target project

In the target project, R_T represents the project scoring matrix. W_T represents the marking matrix in the target project, which is used to distinguish the items in R_T that are not scored from those that are scored. 0 represents that the item is not scored, and 1 represents that the item is scored. The proposed algorithm mainly performs two tasks in the target project. Firstly, it analyses the correlation between the target project and the auxiliary project, and calculates the balance parameters in the UST model; Secondly, the UST model is used to calculate the interest similarity of users in the target project to help users obtain recommended projects. In the auxiliary project, the value of α determines the migration degree of the similarity between the model and user interests. The stronger the correlation between the target project and the auxiliary project, the closer the similarity between users in different projects is. Therefore, the correlation in different fields determines the value of α. In order to measure the relevance of different items, it is expressed by subspace distance formula (13)

$$dist(U_T, U_A) = \sqrt{1 - \sigma \min^2 (U_T^T U_A)} \tag{13}$$

In formula (13), σ represents singular value. U_T and U_A respectively represent the subspaces generated by the scoring matrix of the target project and the auxiliary project. The generated subspaces must meet the orthogonal constraints. In matrix decomposition, a non orthogonal user characteristic matrix is obtained, so it is necessary to orthogonalise the obtained matrix. QR decomposition technology is a common way of matrix decomposition, which can decompose a matrix into a column orthogonal matrix and an upper triangular matrix (Sun et al., 2021; Li et al., 2021; Lucero, 2022). The specific expression is shown in formula (14).

$$U = Q \times T \tag{14}$$

In formula (14), U represents the non-orthogonal matrix obtained during matrix decomposition; Q represents a column orthogonal matrix; T represents the upper triangular matrix. The smaller the value of the subspace formed by scoring evidence, the stronger the correlation of the comparison items, and the greater the value of the migration degree factor; similarly, the larger the value of the subspace, the weaker the

correlation of the comparison items, and the smaller the value of the migration degree factor. Therefore, the migration degree α is calculated as shown in formula (15).

$$\alpha = 1 - dist(U_T, U_A) \tag{15}$$

The final task of the target project is to generate a recommended list suitable for users, so it is necessary to obtain more accurate user interest similarity as far as possible. The method of obtaining interest similarity in the target project is similar to that used in the auxiliary project. The similarity calculation formula is shown in formula (16).

$$T\_sim(u_i, u_j) = \frac{\sum_{k \in I}(r_{i,k} - \overline{r_i})(r_{j,k} - \overline{r_j})}{\sqrt{\sum_{k \in I}(r_{i,k} - \overline{r_i})^2}\sqrt{\sum_{k \in I}(r_{j,k} - \overline{r_j})^2}} \tag{16}$$

According to the above analysis, the specific steps of UST-CF algorithm are as follows. First, enter the scoring matrices of auxiliary items R_A and target items R_T, and the number of adjacent users is k; fill in the missing scoring matrix through the ALS algorithm, and then obtain the prediction scoring matrix Z_A in the auxiliary items; in the matrix filling process, calculate the similarity $A\_sim(u_i, u_j)$ of the common users of the auxiliary project through formula (12); then combine formula (13) with formula (14) to calculate the migration balance parameter α in the model; formula (15) can obtain the interest similarity $T\_sim(u_i, u_j)$ of the original users of the target project. Combining the common user similarity $U\_sim(u_i, u_j)$ obtained by formula (2), the final user similarity of the target project can be obtained. Formula (5) can be used to calculate the prediction scoring matrix Z_A, and finally obtain the recommended list of the target project. The specific flow of UST-CF algorithm is shown in Figure 3.

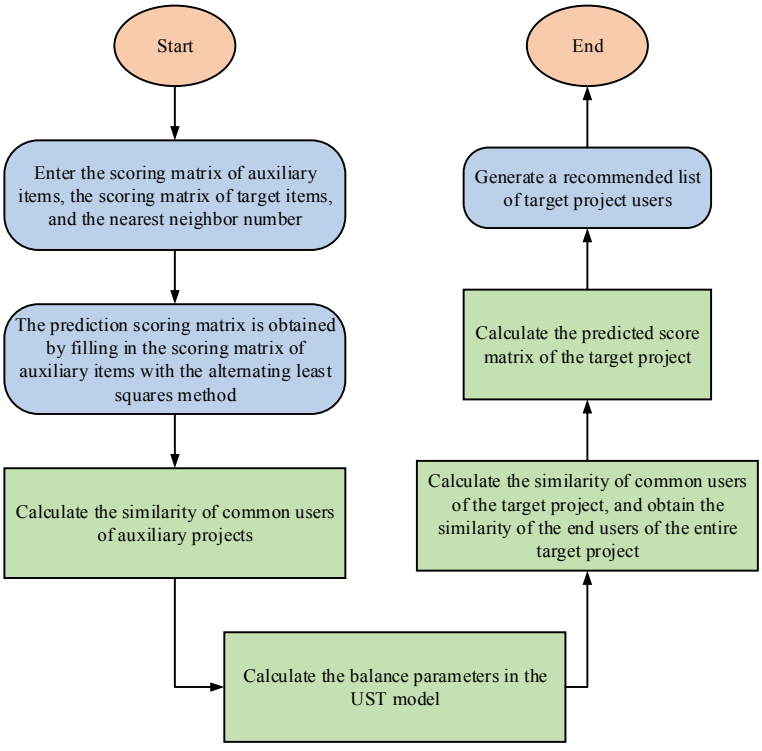
The evaluation index plays an important role in the recommendation system. Through the evaluation index, the deviation between the predicted value and the real value of the model can be analysed to reflect the performance of the recommendation model. The model performance can be judged according to the prediction accuracy, and the prediction accuracy evaluation indicators mainly include mean absolute error (MAE) and root mean squared error (RMSE). The smaller the values of these two indicators, the smaller the prediction error of the model, and the higher the accuracy. The calculation method of MAE is shown in equation (17).

$$MAE = \frac{\sum_{(i,j) \in T_E} \left| p_{i,j} - r_{i,j} \right|}{\left| T_E \right|} \tag{17}$$

In formula (17), T_E represents the test set; $p_{i,j}$ represents the predicted score; $r_{i,j}$ represents the actual score. Similarly, the calculation method of RMSE is shown in formula (18).

$$RMSE = \sqrt{\frac{\sum_{(i,j) \in T_E} (p_{i,j} - r_{i,j})^2}{\left| T_E \right|}} \tag{18}$$

Figure 3 Specific flow chart of UST-CF model generating recommendation list (see online version for colours)

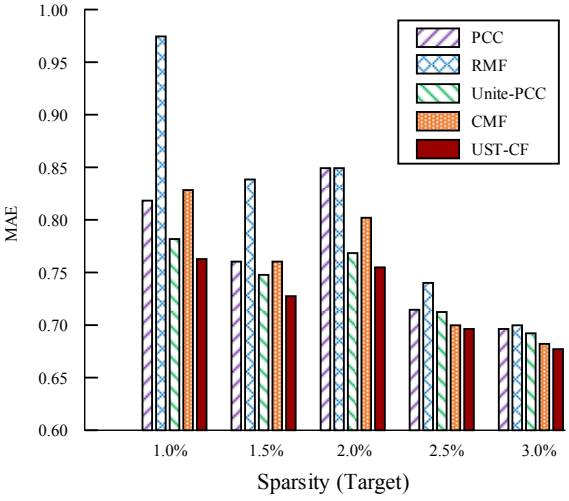


4 Performance analysis of collaborative filtering recommendation algorithm based on user interest similarity migration

4.1 Analysis of recommendation performance under the same user space and the same scoring form

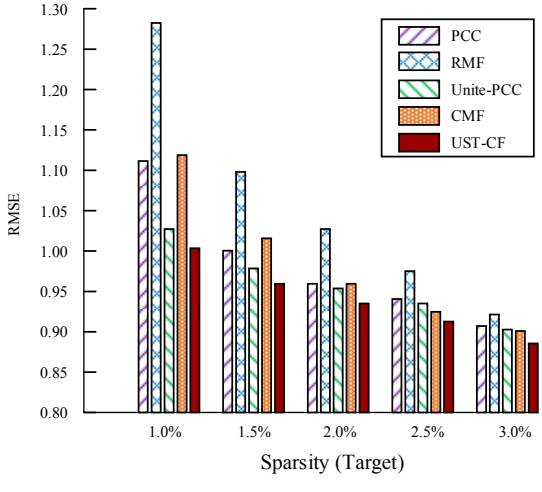
The research verifies the MAE and RMSE index performance of the model by designing different experimental schemes. The remaining five different algorithms are used for comparison with the UST-CF algorithm proposed in the study, including Pearson Correlation Coefficients (PCC), Unite Pearson Correlation Coefficients (Unite PCC), regulated matrix factorisation (RMF), collaborative matrix factorisation (CMF) and User Simulation Transfer unInput (UST unInput). In the experimental scheme with the same user space and scoring form, the sparsity of auxiliary items is fixed at 5%. In the experiment, 500 users are selected to score 1000 learning behaviors, and their 500 learning behaviors are used as the data of auxiliary items, while the remaining 500 learning behaviors are used as the data of target items. In the target project, each user evaluates at least 20 learning behavior styles, and tests them according to different sparsity in the target project. MAE results of different algorithms in target projects with different sparsity are shown in Figure 4.

Figure 4 MAE results of different algorithms in target projects with different sparsity (see online version for colours)



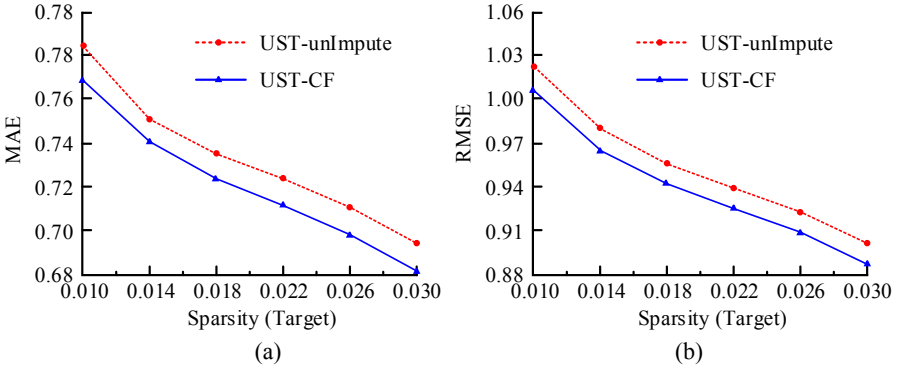
In Figure 4, the sparsity in the target project is divided into five groups, increasing from 0.5% to 3.0%. In the same algorithm, as the proportion of user evaluations increases, the MAE score of the algorithm decreases, indicating that the recommendation accuracy of the algorithm is improved. In the same accuracy, the recommendation effect of using the transfer learning method is obviously better than that of not using the transfer learning method. The MAE values of PCC and RMF algorithms are higher than those of Unite PCC, CMF and UST-CF. The UST-CF algorithm proposed in the study has the lowest MAE value whether in the comparison between models or in the performance comparison of sparsity. When the target item sparsity is 3.0%, the MAE value of the UST-CF algorithm proposed in the study is 0.67. The RMSE results of different algorithms in target projects with different sparsity are shown in Figure 5.

Figure 5 RMSE results of different algorithms in target projects with different sparsity (see online version for colours)



In Figure 5, in the same sparse environment, the RMSE value of the migration learning algorithm is lower than that of the without migration learning algorithm, and the UST-CF algorithm has the best performance. In the same algorithm, the lower the sparsity of the target item, the higher the recommendation performance of different algorithms. When the sparsity of UST-CF algorithm is 3.0%, the RMSE value is 0.88. From the comprehensive analysis of Figures 4 and 5, the lower the sparsity of the target item, the greater the gap between the MAE values of other algorithms and the UST-CF algorithm. It shows that this algorithm has obvious advantages in the case of low sparsity, and has a better mitigation effect on data sparsity. To verify whether the missing matrix in the filling auxiliary items is conducive to improving the final recommendation effect of the model, the experiment adds the UST unInput algorithm without filling for comparison. The comparison results are shown in Figure 6.

Figure 6 Comparison of recommendation effect between filled missing matrix and unfilled missing matrix: (a) comparison of MAE results between UST-CF algorithm and UST unInputsuanfa and (b) comparison of RMSE results between UST-CF algorithm and UST unInputsuanfa (see online version for colours)



In Figure 6(a), the MAE of the model without missing matrix is 0.782 when the sparsity is high; when the sparsity is low, the MAE is 0.697. The MAE of the model filled with missing matrix is 0.769 when the sparsity is low; when the sparsity is low, the MAE is 0.681. In Figure 6(b), when the sparsity is 0.01, the RMSE values of the filled algorithm and the unfilled algorithm are 1.002 and 1.022 respectively; when the sparsity is 0.03, the RMSE values of the filled algorithm and the unfilled algorithm are 0.888 and 0.900 respectively. The results show that the algorithm recommendation accuracy is significantly improved after the missing matrix is filled.

4.2 Analysis of recommendation performance under the same user space and different scoring forms

The experimental scheme adopts the same scoring data as above. Convert the scoring data in auxiliary items, mark the scoring items with a score of 5 or more as like, and record them as 1; items scored less than 5 points are marked as disliked and recorded as 0. The Unite PCC model is required to have a common scoring form for the project, which is not applicable to the scheme. MAE results of other models in this scheme are shown in Table 2.

Table 2 Comparison of MAE results of different models in different sparsity in Scheme II

Algorithm	*Sparsity*				
	1% (MAE)	*1.5% (MAE)*	*2% (MAE)*	*2.5% (MAE)*	*3.0.% (MAE)*
PCC	0.8822	0.8412	0.8167	0.7877	0.7485
RMF	1.1068	0.9498	0.8745	0.8189	0.7769
CMF	1.0469	0.8863	0.8173	0.7604	0.7395
UST-unImpute	0.8872	0.8020	0.7620	0.7409	0.7191
UST-CF	0.8711	0.7913	0.7482	0.7303	0.7095

In Table 2, due to different scoring forms, the recommendation accuracy of PCC algorithm has not improved significantly. When the sparsity is 1.0%, its MAE score is 0.8822; when the sparsity is 3.0%, the MAE score is 0.7485. The proposed UST model has significantly improved the recommendation accuracy in both unfilled and filled forms. When the sparsity is 3.0%, the MAE score of the unfilled UST model is 0.7191; the MAE score of the filled UST model is 0.7095. The RMSE results of other models in this scheme are shown in Table 3.

Table 3 Comparison of RMSE results of different models in different sparsity in Scheme II

Algorithm	*Sparsity*				
	1% (RMSE)	*1.5% (RMSE)*	*2% (RMSE)*	*2.5% (RMSE)*	*3.0.% (RMSE)*
PCC	1.1801	1.1014	1.0599	1.0246	0.9739
RMF	1.4175	1.2234	1.1281	0.0571	1.0081
CMF	1.3430	1.1513	1.0620	0.9850	0.9582
UST-unImpute	1.1809	1.0527	0.9882	0.9620	0.9346
UST-CF	1.1639	1.0377	0.9714	0.9488	0.9207

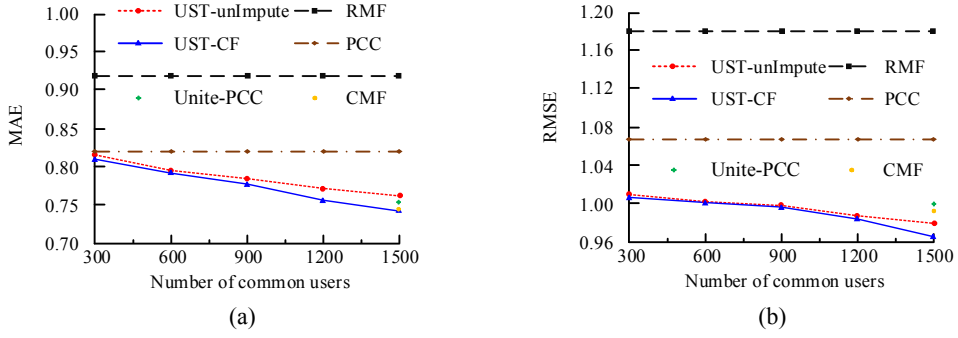
In Table 3, the RMSE scores of the UST model proposed in the study drops below 0.95 when the sparsity is 3.0%. The MAE score of the unfilled UST model is 0.9346, and the MAE score of the filled UST model is 0.9207. Through comprehensive analysis and comparison between Tables 2 and 3, when different scoring forms are used in auxiliary projects and target projects, the UST model proposed in the study has better adaptability and recommendation effect in practical application.

4.3 Analysis of recommendation performance with different user spaces and the same scoring form

In this experimental scheme, the experiment is divided into different groups according to the number of common users of the auxiliary project and the target project. The total number of users is increased to 1500, and the first group is 300. The number of common users of each group is increased by 300 from the previous group, which is divided into five groups. The fifth group is 1500. In the target project, each user should evaluate at least 20 learning behaviors and keep the sparsity of the project at 1.0%. Because Unite PCC and CMF models need to use common features to decompose the scoring matrix, the model has no decomposition effect in some common users. Therefore, the two models

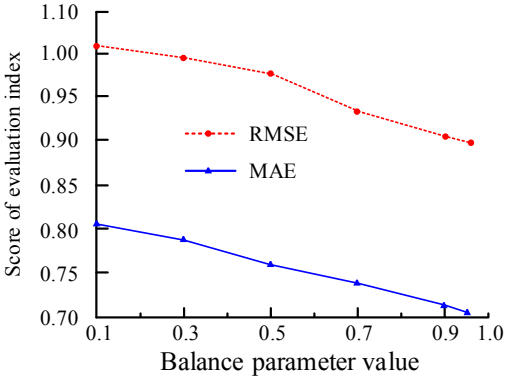
only present the final comparison results. MAE results of different models in Scheme 3 are shown in Figure 7.

Figure 7 MAE results of different models in Scheme 3: (a) MAE values of different models in Scheme III and (b) RMSE values of different models in Scheme III (see online version for colours)



In Figure 7 (a), the MAE values of PCC and RMF algorithm do not change. Therefore, in the environment of some common users, the recommendation effect of PCC algorithm and RMF algorithm needs to be improved. However, the MAE value of the UST model in different user numbers is changing. With the increase of the number of common users, the MAE value is decreasing. Thus, the number of common users has a certain impact on the recommendation effect of the UST model. In Figure 7(b), the RMSE scores of PCC and RMF algorithms change in different user environments. In the UST model, the number of users is inversely proportional to the RMSE score. The more user information, the more effective the improvement of model accuracy. Based on the analysis of the results in Figure 7, the UST-CF algorithm makes more thorough use of the user information of auxiliary projects. Therefore, the recommendation accuracy of the model can still be improved when there are fewer common users. The performance of UST-CF algorithm is verified. The balance parameters in the algorithm also have an impact on the prediction effect. Therefore, the recommended performance effects of UST-CF in different balance parameters are shown in Figure 8.

Figure 8 Recommended performance effect of UST-CF in different balance parameters



In Figure 8, the larger the balance parameter, the lower the MAE and RMSE scores of the UST-CF model. When the balance parameter is 0.95, the MAE and RMSE scores of the model are 0.7066 and 0.8978 respectively. In the experiment of adjusting the balance parameter, the higher the similarity between the transfer learning of auxiliary items and target items, the better the recommendation performance of the model.

5 Conclusion

In the network TCFL, the network teaching platform provides users with teaching content and teaching methods that are difficult to satisfy users. Therefore, in order to make the TCFL develop in a higher quality, the research uses the model to transfer the interest similarity, so as to improve the recommendation effect of the recommendation model. The method of user similarity is applied to rating users. This method is used for solving the sparsity problem of collaborative filtering technology, and can obtain suggestions from similar users' information. In addition, this method shows good performance in a variety of datasets. Especially when the data sparsity is high, it can still play a role in recommendation (Hu and Schwartz, 2022; Zhu et al., 2021). SVD method calculates the maximum singular value and ahead singular value of sparse matrix to fill the matrix. This method is widely used in image restoration and recommendation systems (Ortal and Edahiro, 2020; Kai et al., 2021). The UST-CF model is proposed using the above methods, and different non migration learning algorithms and migration learning algorithms are compared with them. The experimental results show that the overall performance of the model using transfer learning is obviously better than that of the model using non transfer learning; in different sparsity environments, the lower the sparsity, the better the migration effect of user interest similarity; the transfer learning model still has strong adaptability in the environment with high sparsity. By filling in the missing matrix in the model and calculating the similarity between different items, the recommendation performance of the model is improved. Three schemes are designed to verify the performance of the UST-CF model proposed in the study. The analysis results show that the performance of the UST-CF model in different experimental schemes is higher than that of other models. In Scheme I, the model has a significant gap with other models in an environment with the sparsity of 1.0%; when the sparsity is 3.0%, the MAE and RMSE scores are 0.67 and 0.88 respectively; in Scheme II, the model has better adaptability to different scoring forms. When the sparsity is 3.0%, the MAE and RMSE scores are 0.7095 and 0.9346 respectively; in Scheme 3, the UST-CF model performs better in using the user information of auxiliary projects, and has good recommendation effect when the number of common users is small. It is a relatively new method to combine the innovative way of teaching CFL with the network recommendation model. The UST-CF model proposed in the study effectively alleviates the problem of data sparsity in the target field, improves the intelligence and recommendation quality of the model, makes the UST model have extensive applicability and strong stability, and provides a new research direction for the development of recommendation systems. However, there are still some deficiencies in the research. The sample data used in the study is small, and the same weight value is applied to transfer learning. Therefore, the follow-up research also aims at improving the problems. Subsequent research can expand the sample data to further optimise the recommendation effect, and optimise the weight

of transfer learning to make the model recommendation effect more reasonable and scientific.

References

Ajaegbu, C. (2021) 'An optimized item-based collaborative filtering algorithm', *Journal of Ambient Intelligence and Humanized Computing*, Vol. 12, No. 12, pp.10629–10636.

Attaran, M. and Yishuai, H. (2018) 'Teacher education curriculum for teaching Chinese as a foreign language', *MOJES: Malaysian Online Journal of Educational Sciences*, Vol. 3, No. 1, pp.34–43.

Barbosa, B., Saura, J.R. and Bennett, D. (2022) 'How do entrepreneurs perform digital marketing across the customer journey? A review and discussion of the main uses', *The Journal of Technology Transfer*, pp.1–35.

Chen, J., Li, K., Rong, H., Bilal, K., Yang, N. and Li, K. (2018) *A Disease Diagnosis and Treatment Recommendation System Based on Big Data Mining and Cloud Computing*, arXiv e-prints, Vol. 435, pp.124–149.

Du, Q., Youngentob, S. and Middleton, F. (2019) 'Fetal alcohol exposure in rats leads to long-term changes in DNA methylation patterns of taste-related genes in postnatal tongue', *European Neuropsychopharmacology*, Vol. 29, pp.S917–S918.

Envelope, T. (2021) 'Research on automatic user identification system of leaked electricity based on data mining technology', *Energy Reports*, Vol. 7, pp.1092–1100.

Fan, H., Wu, K., Parvin, H., Beigi, A. and Pho, K. (2021) 'A hybrid recommender system using KNN and clustering', *International Journal of Information Technology and Decision Making*, Vol. 20, No. 2, pp.553–596.

Gong, Y., Lyu, B. and Gao, X. (2018) 'Research on teaching Chinese as a second or foreign language in and outside mainland China: a bibliometric analysis', *The Asia-Pacific Education Researcher*, Vol. 27, No. 4, pp.277–289.

Grl, A., Cem, B., Al, C. and Gz, A. (2020) 'Applying landmarks to enhance memory-based collaborative filtering', *Information Sciences*, Vol. 513, pp.412–428.

Han, X., Wang, Z. and Xu, H.J. (2021) 'Time-weighted collaborative filtering algorithm based on improved mini batch K-means clustering', *Advances in Science and Technology*, Trans Tech Publications Ltd., Vol. 105, pp.309–317.

Hu, V.W. and Schwartz, D.T. (2022) 'Low-error estimation of half-cell thermodynamic parameters from whole-cell Li-ion battery experiments: physics-based model formulation, experimental demonstration, and an open software tool', *Journal of The Electrochemical Society*, Vol. 169, No. 3, pp.1–16.

Iwendi, C., Ibeke, E., Eggoni, H., Velagala, S. and Srivastava, G. (2022) 'Pointer-based item-to-item collaborative filtering recommendation system using a machine learning model', *International Journal of Information Technology and Decision Making*, Vol. 21, No. 01, pp.463–484.

Jin, X. and Hu, H. (2022) 'Research and implementation of smart energy investment and financing system design based on energy mega data mining', *Energy Reports*, Vol. 8, pp.1226–1235.

Kai, X.A., Ying, Z.B. and Zhi, X. (2021) 'Iterative rank-one matrix completion via singular value decomposition and nuclear norm regularization', *Information Sciences*, Vol. 578, pp.574–591.

Li, J.T. and Tong, F. (2019) 'Multimedia-assisted self-learning materials: the benefits of E-flashcards for vocabulary learning in Chinese as a foreign language', *Reading and Writing*, Vol. 32, No. 5, pp.1175–1195.

Li, L. (2021) 'A case study on the pattern of classroom discourse in teaching Chinese as a foreign language', *Journal of Contemporary Educational Research*, Vol. 5, No. 10, pp.172–177.

Li, P., Wang, H., Li, X. and Zhang, C. (2021) 'An image denoising algorithm based on adaptive clustering and singular value decomposition', *IET Image Processing*, Vol. 15, No. 3, pp.598–614.

Li, Z., Zhang, Y., Ming, L., Guo, J. and Katsikis, V.N. (2021) 'Real-domain QR decomposition models employing zeroing neural network and time-discretization formulas for time-varying matrices', *Neurocomputing*, Vol. 448, No. 14, pp.217–227.

Lucero, J.C. (2022) 'Identification of independent patterns of COVID-19 mortality in Brazil by a functional QR decomposition', *Biophysical Reviews and Letters*, Vol. 17, No. 1, pp.33–41.

Mosavi, N.S., Freitas, F., Pires, R., Rodrigues, C., Silva, I., Santos, M. and Novais, P. (2022) 'Intelligent energy management using data mining techniques at Bosch car multimedia Portugal facilities', *Procedia Computer Science*, Vol. 201, pp.503–510.

Na, L., Li, M.X., Qiu, H.Y. and Su, H.L. (2021) 'A hybrid user-based collaborative filtering algorithm with topic model', *Applied Intelligence*, Vol. 51, No. 11, pp.7946–7959.

Nguyen, N.T.T. (2021) 'A review of the effects of media on foreign language vocabulary acquisition', *International Journal of TESOL and Education*, Vol. 1, No. 1, pp.30–37.

Ortal, P. and Edahiro, M. (2020) 'Switching hybrid method based on user similarity and global statistics for collaborative filtering', *IEEE Access*, Vol. 8, pp.213401–213415.

Ribeiro-Navarrete, S., Saura, J.R. and Palacios-Marqués, D. (2021) 'Towards a new era of mass data collection: assessing pandemic surveillance technologies to preserve user privacy', *Technological Forecasting and Social Change*, Vol. 167, No. 1, pp.120681.1–120681.14.

Saura, J.R., Palacios-Marqués, D. and Ribeiro-Soriano, D. (2023) 'Exploring the boundaries of open innovation: evidence from social media mining', *Technovation*, Vol. 119, pp.102447–102447.

Saura, J.R., Ribeiro-Soriano, D. and D., Palacios-Marqués (2021) 'Setting B2B digital marketing in artificial intelligence-based CRMs: a review and directions for future research', *Industrial Marketing Management*, Vol. 98, No. 4, pp.161–178.

Schwalbe-koda, D., Santiago-reyes, O.A., Corma, A., Roman-leshkov, Y., Moliner, M. and Gomez-bombarelli, R. (2022) 'Repurposing templates for zeolite synthesis from simulations and data mining', *Chemistry of Materials: A Publication of the American Chemistry Society*, Vol. 34, No. 12, pp.5366–5376.

Sun, L., Wu, B. and Ye, T. (2021) 'Design and VLSI implementation of a sorted MMSE QR decomposition for 4 × 4 MIMO detectors', *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, Vol. E104/A, No. 4, pp.762–767.

Xing, Y., He, W., Cao, G. and Li, Y. (2022) 'Using data mining to track the information spreading on social media about the COVID-19 outbreak', *The Electronic Library: The International Journal for Minicomputer, Microcomputer, and Software Applications in Libraries*, Vol. 40, Nos. 1–2, pp.63–82.

Yang, J., Yin, C. and Wang, W. (2018) 'Flipping the classroom in teaching Chinese as a foreign language', *Language Learning and Technology*, Vol. 22, No. 1, pp.16–26.

Yu, A. and Geng, G. (2020) 'Introducing Chinese Hanzi at the beginning of teaching Chinese as a Foreign Language (CFL)', *Global Chinese*, Vol. 6, No. 2, pp.359–380.

Zeng, Y. and Jiang, W. (2021) 'Barriers to technology integration into teaching Chinese as a foreign language: a case study of Australian secondary schools', *World Journal of Education*, Vol. 11, No. 5, pp.17–30.

Zhang, S. and lynch, R. (2021) 'The relationship of self-efficacy for Chinese as a foreign language oral skills and the use of indirect language learning strategies for learning Chinese as a foreign language with oral skills achievement of grade 6 students in Chinese as a foreign language class at Satit Prasarnmit Elementary School in Bangkok, Thailand', *Scholar: Human Sciences*, Vol. 13, No. 1, pp.105–120.

Zhu, X., Duan, S.B., Li, Z.L., Zhao, W., Wu, H., Leng, P., Gao, M.F. and Zhou, X.M. (2021) 'Retrieval of land surface temperature with topographic effect correction from Landsat 8 thermal infrared data in mountainous areas', *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 59, No. 8, pp.6674–6687.