International Journal for Housing Science and Its Applications
Publish August 4, 2025. Volume 46, Issue 3 Pages 2132-2142

https://doi.org/10.70517/ijhsa463169

A Study of Semantic Similarity and Language Selection Strategies Based on Dynamic Planning Model in English Cultural Communication

Zhi Zhang[1,*]
[1] Foreign Language Department, Lyu Liang University, Lv Liang, Shanxi, 033001, China
Corresponding authors: (e-mail: zhangzhikaoyan@163.com).

Abstract This study proposes an intelligent service recommendation system that integrates dynamic planning model and improved label propagation algorithm (OCDSLP). Aiming at the limitations of traditional methods in semantic association mining and cross-cultural interaction adaptation, a three-tier system framework based on B/S architecture is firstly constructed to realize the separation of multi-role authority management and layered business logic. The OCDSLP algorithm is then proposed to optimize the community discovery process by fusing network topology similarity and semantic similarity. The state compression dynamic planning model is introduced to filter the optimal service combination segments. Experiments show that the OCDSLP algorithm achieves semantic matching accuracy of 85.1% and 84.3% on Usedcar DB and IMDB datasets. In the English cultural communication platform application, the proposed model performs optimally over the six control models under different number of Top-K neighbors. In the MAE dimension when the number of neighbors is 20, the error value reaches the lowest, only 18.891, and it is about 15.3% lower than the next best method NMF (RMSE=62.986) at the RMSE dimension threshold of 20. The results of the questionnaire survey show that users recognize more than 80% of the system functionality and experience.

Index Terms dynamic programming model, semantic similarity, OCDSLP algorithm, English cultural communication

I. Introduction

With the deepening of economic globalization, cross-cultural communication is becoming more and more frequent, and English, as one of the global languages, plays an important role in cross-cultural communication. English cultural communication can not only improve the individual's English proficiency and make him/her better integrated into the international society, but also spread the national culture and lead the fashion trend in the cultural world [1]. However, due to the differences in cultural backgrounds, ways of thinking, semantic understanding, language strategies, etc., inappropriate language expressions may lead to communication barriers or even trigger misunderstandings [2]-[4]. In the fields of academic communication, international trade, medical cooperation, and scientific and technological development, the percentage of misunderstanding under semantic similarity that leads to communication barriers is as high as 25% or more. In addition, language strategies are selected based on contextual analysis and communication purposes, but the variability of implicit features of contexts leads to unsatisfactory communication effects [5], [6]. How to avoid semantic similarity misunderstanding and appropriately use English language selection strategies in cross-cultural communication to improve the efficiency of communication has become a problem worth exploring.

In semantic similarity, researchers focus on the analysis of the language words themselves, and literature [7] introduces to the fact that in the study of English text similarity, the accuracy of the method based on corpus, feature description, and word embedding is better with the feature-cutting method. Literature [8] utilized the knowledge base of Chinese Dictionary to explain the association between lexical concepts, and at the same time united the non-relational databases to design different equations to calculate the semantic similarity between English and Chinese. Literature [9] utilized Word2Vec to convert English vocabulary into vectors, and then carried out English vocabulary similarity calculation by cosine similarity calculation method. Literature [10] used text embedding for vector transformation of concept and category idiosyncrasies, combined with k-kernel method of semantic perception, concept and category graph design, and hybrid weighting method of feature fusion model to obtain the semantic similarity computation strategy. Literature [11] used Transformer's bidirectional encoder representation and support vector machine to represent and categorize the semantics of English text, respectively, to achieve semantic similarity assessment. However, these methods do not consider the semantic similarity in different cultural contexts, and the models and methods are poorly adapted to cultural misunderstanding elimination.

As for the research on language strategies, literature [12] mentions that among the polite communication strategies in cross-cultural scenarios, Asian cultures express themselves in indirect and metaphorical approaches, European cultures express themselves in direct and humorous approaches, and African cultures express themselves in indirect and storytelling approaches. However, in intelligent systems, translations and other platforms, the procedural misunderstanding of fuzzy cultures and cultural contexts results in the lack of correct language strategy selection, which puts cultural communication in diagnosis and treatment, business negotiation and other scenarios at a disadvantage. And the dynamic programming model is an optimization method for solving multi-stage decision-making problems, which divides the problem into several stages, takes an optimal decision at each stage, and obtains the optimal solution of the problem by recursion, which is advantageous in culturally interpretable and cross-cultural speech dialogues [13], [14].

In this paper, we first design a B/S architecture system to support fine-grained permission management of six types of user roles to build an English cultural communication system. By constructing a multi-dimensional semantic similarity computation framework and a state compression dynamic planning model, the optimal retrieval of service combination fragments and the adaptation of English resources are realized. UsedcarDB and IMDB are selected as experimental datasets to evaluate the performance level of OCDSLP algorithm. MAE and RMSE are used as evaluation metrics to examine the application effect of the proposed model. Based on the user trial evaluation, the effectiveness of the proposed system is examined.

II. Combining OCDSLP algorithm and dynamic programming model for English cultural communication system

Under the background of accelerated globalization, English cultural communication, as the core carrier of cross-cultural communication, has a direct impact on the effectiveness of language learning and cultural dissemination in terms of its efficiency and quality. However, existing English learning platforms generally face two major challenges. First, it is difficult to accurately model the dynamic correlation between users' semantic preferences and cultural background, resulting in insufficient cultural adaptability of the recommendation system; second, there is computational redundancy in the optimization of service combinations under massive interaction data, which affects the real-time response capability. Traditional methods rely on a single similarity metric or greedy algorithm, which is easy to fall into local optimization and cannot effectively integrate topological structure and semantic information.

Aiming at the above problems, this paper proposes a solution integrating dynamic planning model and improved label propagation algorithm (OCDSLP), which aims at solving the semantic gap problem in English cultural communication scenarios.

II. A.Analysis and Design of English Cultural Communication System

II. A. 1) Overall architectural design

This system adopts the most widely used three-tier architecture (B/S architecture), which puts the data storage layer in the server as the database server; puts the presentation layer in the client as the user interface; and separates the business logic layer from the client, and the business logic is realized by the part of the application server (also known as the Web server), and the overall architecture of the system is shown in Figure 1.

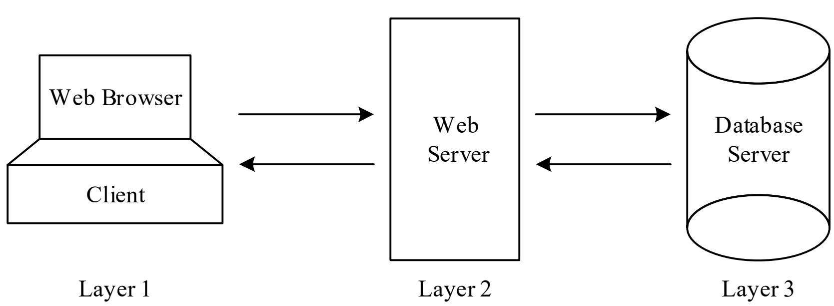


Figure 1: Overall Architecture of the system

II. A. 2) Functional analysis

There are six roles in this system, which are Administrator, Teacher, Academic Committee-cum-Group Leader, Academic Committee, Group Leader, and Ordinary Student, and each role corresponds to one kind of permission. In the design, the practice of user-associated roles and role-associated permissions is used, and the permissions are controlled at the button level in order to reduce the number of JSP pages. The system implementation of permission assignment and role association is shown in Figure 2.

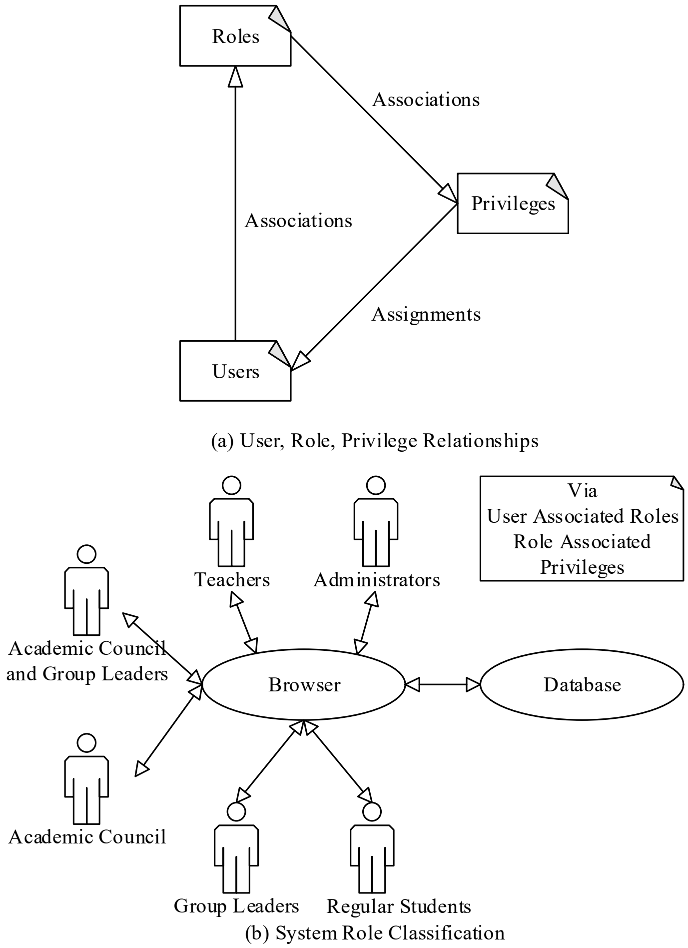


Figure 2: Permission allocation and role association of the system

After a user is matched with a role, the corresponding permissions are granted. According to the requirements, the corresponding permissions for different roles are shown in Table 1.

Table 1: Roles and Corresponding Permissions

Role	Permission
Public authority	Forget passwords,change passwords,manage personal information,and access login pages
Ordinary student	Manage personal documents and plans,download teaching materials,view project and group information,and evaluate courses
Group leader authority	General student permissions, selection of group members,management of group reports,management of group plans,task allocation for group members
The authority of the academic committee	The authority for ordinary students,the designation of class leaders,and the management of students' group selection information
Academic committee member and group leader	The authority of the group leader and that of the academic committee member
Teacher	Assign study committee members and group leaders,manage teams,manage teaching materials,and review course evaluations
Administrator	Teacher permissions,manage teachers/students,initialize user password management dictionary table,view system logs

II. B.Label pre-assignment based on OCDSLP algorithm

Based on the traditional label propagation algorithm COPPRA, this paper proposes the OCDSLP method to improve COPPRA by utilizing user semantic information and network topology information, and the principle and framework are shown in Fig. 3. In Fig. 3, the OCDSLP method consists of four modules: data preprocessing, label pre-assignment, label propagation and community discovery. In the first module, it mainly preprocesses the social network dataset and mines the network topology information and users' attribute information; in the second module, it mines the users' semantic similarity and topology similarity and fuses the two to generate the integrated similarity and combines them with the node activity level for node label pre-assignment; in the third module, it integrates the semantic similarity and topology similarity of nodes and improves the label attribution coefficients and propose label importance and label activity for label optimization and updating; finally, in the fourth module, the final community is divided based on the labels owned by the nodes.

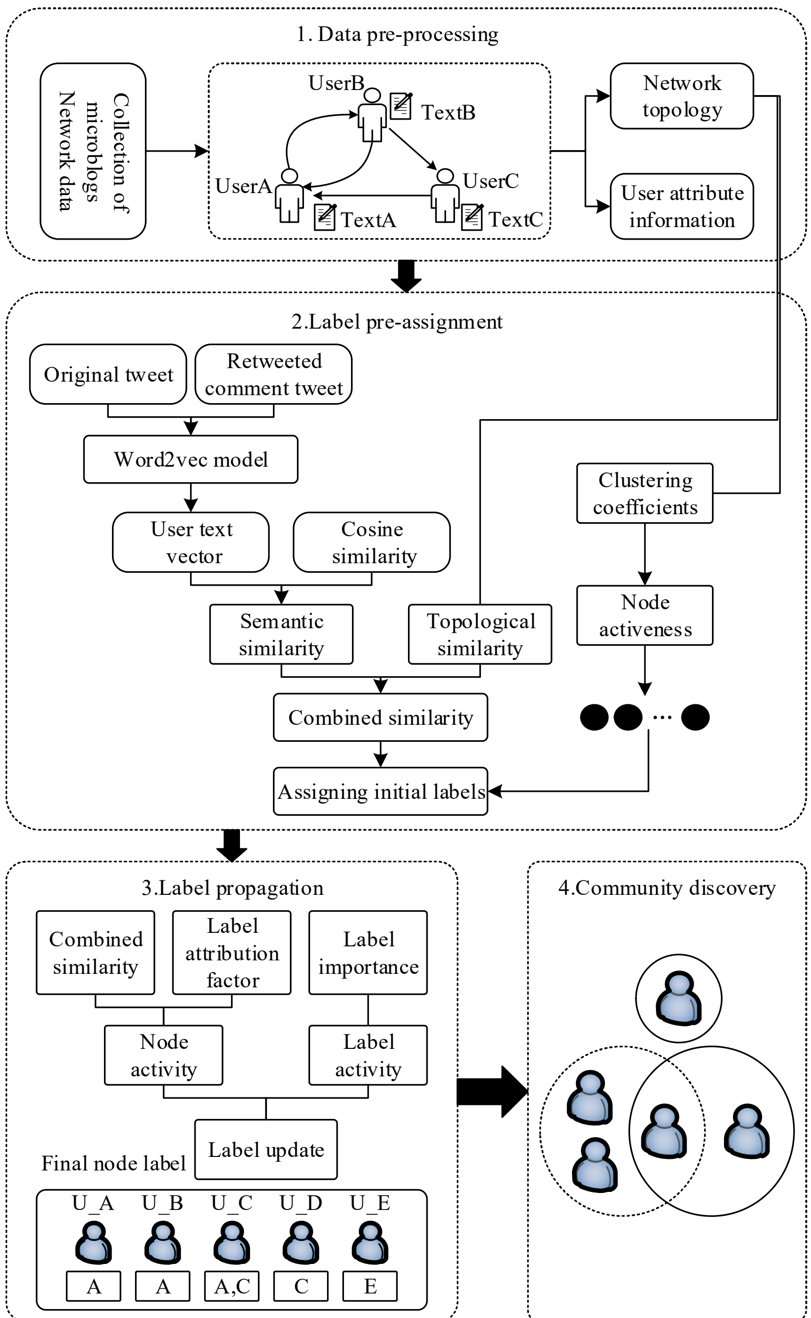


Figure 3: The principle and framework of the OCDSLP method

II. B. 1) Topological similarity

The number of common neighbor nodes of a node reflects the correlation between nodes to a certain extent, the more common neighbors of a node indicates that the two nodes are more closely connected in the network and the higher the similarity. Therefore, this paper defines the topological similarity of nodes based on Jaccard coefficient, as shown in equation (1).

$$SIM\_t_{u,v} = \frac{|N(u) \cap N(v)| + 1}{|N(v) \cup N(u)|} \tag{1}$$

where: $N(u)$, $N(v)$ denotes the number of neighboring nodes of node u and v, and the numerator plus 1 is to prevent the value of $SIM\_t_{u,v}$ from being 0.

II. B. 2) Semantic similarity

The higher semantic similarity of users indicates that users are more likely to share the same interest preferences and have a higher probability of being in the same community, and conversely users share fewer common interest preferences and have a lower probability of belonging to the same community.

In the OCDSLP algorithm, firstly, the deactivated words are removed and the key words that identify the user's interests are retained, i.e., the user's semantic information; then, the word vectors of each word are generated based on the Word2Vec language model, and then, the interest vectors of the user, i.e., the user's semantic features, are generated for all word vectors corresponding to each user; and finally, the cosine similarity is utilized to measure the semantic similarity among the users as shown in Eq. (2).) is used to measure the semantic similarity between users, as shown in equation (2).

$$SIM\_s_{u,v} = \frac{\sum_i^n u_i * \sum_i^n v_i}{\sqrt{\sum_i^n u_i^2} * \sqrt{\sum_i^n v_i^2}} \tag{2}$$

where: $SIM\_s_{u,v}$ denotes the semantic similarity between users u and v; u_i and v_i denote user u and v text vectors, respectively.

II. C.Service Recommendation Based on Dynamic Planning Models

The dynamic programming (DP) algorithm is a classical algorithm in computer science. In computer science, mathematics, management science, economics, and bioinformatics, dynamic programming is a method of solving complex problems by decomposing the problem to be solved into a series of simple subproblems, solving them one at a time and storing their results. When the same subproblem arises next time, the results of the previous computation are used directly instead of recomputing its solution, achieving the goal of saving computation time by a simple lookup of the previously computed solution. Of course, this consumes some extra storage space, and the technique of storing the solutions of the subproblems instead of recomputing them is called “memoization”. Dynamic programming algorithms are often used for optimization. Dynamic programming algorithms examine previously solved subproblems and combine their solutions to give the best solution for a given problem. In contrast, greedy algorithms treat the solution as a series of steps and choose a local optimum at each step. Using the greedy algorithm does not always guarantee an optimal solution, whereas the dynamic programming algorithm does, because choosing a local optimum may lead to a bad global solution. One advantage of greedy algorithms over dynamic programming algorithms is that greedy algorithms tend to be faster and simpler. Some greedy algorithms (e.g., Kraskal or Primm for minimum spanning trees). Dynamic planning algorithms and partitioning algorithms also share many similarities in that both solve the problem by breaking it down into smaller subproblems. The difference is that the partitioning algorithm requires the original problem to be decomposed into a number of sub-problems that do not contain a common subproblem, whereas dynamic programming contains a common subproblem.

Dynamic programming generally has to have three properties:

(1) Optimization principle: assuming that the optimal solution of the problem includes the solution of the subproblem is also optimal, it is said that the problem has the optimal substructure, that is, to satisfy the optimization principle;

(2) No posteriority: that is, once a state is determined at a certain stage, it is not affected by decisions made after this state. In other words, the process after a certain state will not affect the state once, only related to the current state;

(3) With overlapping subproblems: i.e., subproblems are not independent of each other, and a subproblem may be used multiple times in the next stage of decision making.

There are many examples of dynamic programming algorithms applied to service combination and recommendation. By using dynamic programming algorithms to retrieve suitable service combination fragments from a derived subgraph, we aim to convert a multi-stage process into a series of single-stage problems. In order to retrieve the optimal service combination fragments quickly, this paper employs a state-compressed dynamic programming algorithm. The specific algorithm is described as follows:

In the algorithm, st is a binary bit of n bits used to indicate the connection between n operations in the service network model. Each binary bit is a label indicating whether the current operation is contained in the service combination fragment. When the binary bit is 1, the current operation is included, and 0 indicates that the operation is not included. Indicates a service combination fragment between n operations in a service network model. Each binary bit represents the label of an operation, e.g., $st = 21$, whose binary can be represented as 10101, the path passes through operations 1, 3, and 5. when $st = 2^n - 1$, n operations are contained in that service combination fragment. ed denotes the corresponding end operation under the path st, and $path_{ed}$ denotes the predecessor node of the end node under the current service combination fragment, which is used to record the order of operations in the final service combination fragment.

Our algorithm through dynamic programming is implemented in the following way, first the maximum value of st is given, then st will iterate through each possible state of st starting from 0, and all possible termination operations in that state are obtained by iteration. When st and ed are given, the state of the next service fragment combination is iterated after adding all new operations. At the beginning of each state, the current path is first checked to see if it contains the operation ed, and if the current state st does not contain the operation ed, then the variance var, the mean $aver$, and the objective function af are computed for the new fragment. When the value of the objective function af is greater than af_{temp}, we continue to check the next operation if the state st contains operation ed until all operations in ed are checked. Meanwhile, the path process is recorded. The update iteration process ends when $st = 2^n - 1$. Select one of the n operations as the end operation for that state, and record the service combination fragment with the largest objective function in the current state. The path order is also recorded and finally the service combination fragment with the largest objective function is recommended to the user.

$$af_{temp} = 1 + aver - 2 \times var \tag{3}$$

The role of dynamic programming algorithm is to retrieve the optimal solution in the service network model, by which we find the optimal service combination segments. As shown in equation (3), $aver$ denotes the mean of the service combination segments. var denotes the variance of the service combination segments. af_{temp} represents the objective function of the service combination segment. The objective function is used to represent the satisfaction of the service combination segments to indicate the quality of the service combination segments. Therefore, we need to evaluate the service portfolio segments from the following two aspects:

(1) A service combination fragment is more reasonable when the likelihood of invocation between operations on the service combination fragment is relatively high. We need that the individual operations in the fragment should have the highest possible semantic similarity to the user's requirements in order to obtain a more accurate service combination fragment, so we use the average value as part of the objective function.

(2) When the difference of the mean value between the individual operations contained in a service combination fragment is small, the fragment will be recommended to the user preferentially, and the difference of the weight of the service combination fragment is measured by the variance. This parameter ensures that each operation in the fragment has a high weight, and the situation that the similarity between two of the service combination fragments is very low while the similarity between other operations is very high can be avoided as much as possible. The service combination segments that fulfill the above two conditions will be recommended to the user.

For example. The current state st runs from 0 to $2^5 - 1 = 31$ when the service combination fragment needs to traverse a complete graph consisting of 5 operations. If the current state is 01000. And if the next operation is not already included in the fragment. It is added to the current service combination fragment. Some new state will be generated at this point. In the next cycle, continue generating a new state through the previous state until all operations are included in the service combination fragment. It is worth noting that when all operations are included, there will be many different kinds of states generated, and each time a different order of operations is added, different results will be generated, so for the final inclusion of all operations, the different cases are distinguished by recording the order of operations in the service combination fragment.

III. Empirical study on the application effect of English cultural communication system based on dynamic planning modeling

III. A. Algorithm validation

The experimental test machine was configured with an IntelP4 processor at 3.2GHz, 8GB of RAM, and Windows 2008 operating system. All algorithms were implemented using C# and SQL. The experimental data is obtained using 2 test datasets: one is 200,000 used car information records randomly selected from Yahoo! Autos website, synthesized into the relational table Usedcar DB. The other is IMDB: this dataset contains 5 relational tables: Actors, Roles, Movies, Diretor_Movie, and Directors tables.

This section tests the accuracy of the proposed OCDSLP algorithm with a user survey. Ten users (PhD students, Masters students and young faculty members) were invited to execute 10 queries each from UsedcarDB and IMDB datasets (1 query per user per dataset). For each query, 30 semantically relevant query results are returned for the tuple according to its satisfaction level of the query (by observing these results, it can be found that some tuples are very similar to each other), and then the user is asked to annotate the top 5 representative and relevant tuples among these results. Then, the OCDSLP algorithm proposed in this paper and the traditional sorting method are used to get the top 5 tuples respectively, based on which the overlap between the results obtained by the two methods and the results labeled by the users is tested, and the higher the overlap, the higher the user satisfaction (i.e., accuracy). The accuracy is calculated according to equation (4):

$$Accuracy = \left(\left|top-5\ tuples\ retrieved \cap relevant\ tuples\right|\right)/5 \tag{4}$$

where the numerator of Eq. (4) is the intersection between the top-5 tuples retrieved by different methods and the top 5 representative relevant tuples labeled by the user for the query. Since the value of k in $top-k$ is set to 5 in this paper, the denominator is 5.

A comparison of the accuracy of the two query result selection methods on the datasets Usedcar DB and IMDB is shown in Figure 4. It can be seen that there is a high overlap between the typical tuples obtained by the OCDSLP algorithm using semantic similarity and the representative tuples labeled by the users, and the accuracy on the Usedcar DB and IMDB datasets reaches 85.1% and 84.3%, respectively (the accuracy is computed by taking the average of the 10 test queries), whereas the selection accuracy of the traditional ranking method is only 52.0% and 56.9%.

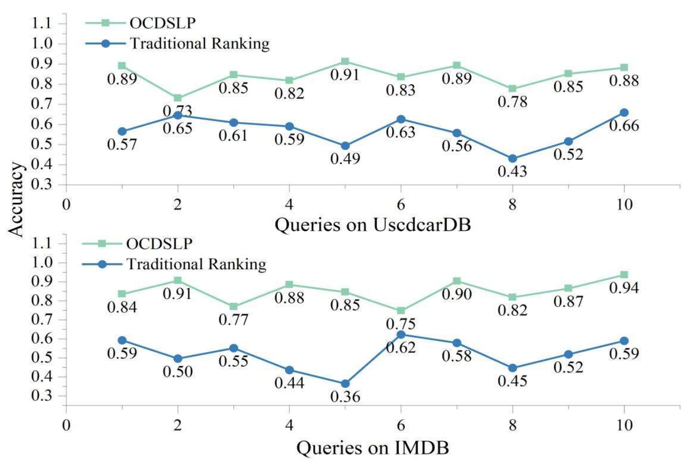


Figure 4: Accuracy comparison

III. B. Application effects

The data in this section comes from the user-generated content of HelloTalk, an international language learning community, from April to June 2024, covering the recorded interactions between English and non-native English speakers. The original data volume is 12,450 entries, and 6,200 valid conversations are retained after cleaning (removing duplicates, non-English texts), each containing user-input text, target language responses, cultural tags, and expert-annotated cultural appropriateness scores. Two criteria, mean absolute error (MAE) and root mean square error (RMSE), are used to measure the accuracy of the model proposed in this paper.

The accuracy of the dynamic planning model with or without the OCDSLP algorithm is first explored, focusing on the comparison of the accuracy of the two methods under different number of neighbors. In order to highlight the performance advantage of the dynamic programming model with OCDSLP algorithm when dealing with sparse datasets, the experiment set the sparsity of the dataset as 50%, and selected the number of neighbors from 10 to 50, with 10 as the interval, for comparison. The experimental results are shown in Fig. 5.

From the results in Fig. 5, it can be observed that the proposed model outperforms the model without label preassignment using the OCDSLP algorithm under 50% data sparsity for each neighbor number setting. Specifically, the proposed model achieves its optimal performance when the number of neighbors is 10, while the control model achieves its optimal performance when the number of neighbors is 50, but the proposed model outperforms the control model even when the number of neighbors is 10. This result shows that the model obtains an improvement in prediction accuracy and has a significant advantage in handling sparse datasets after using the OCDSLP algorithm for label preassignment.

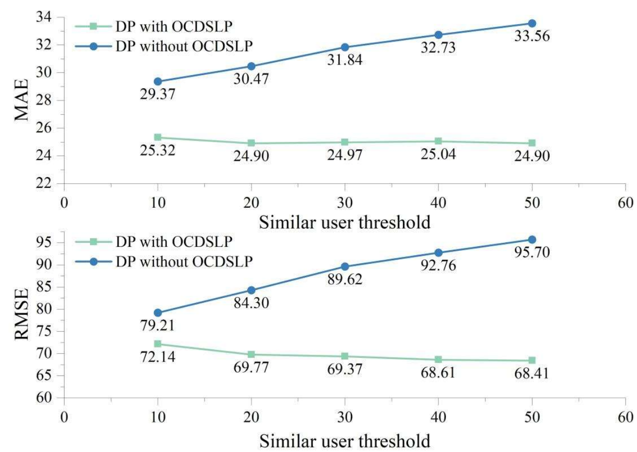
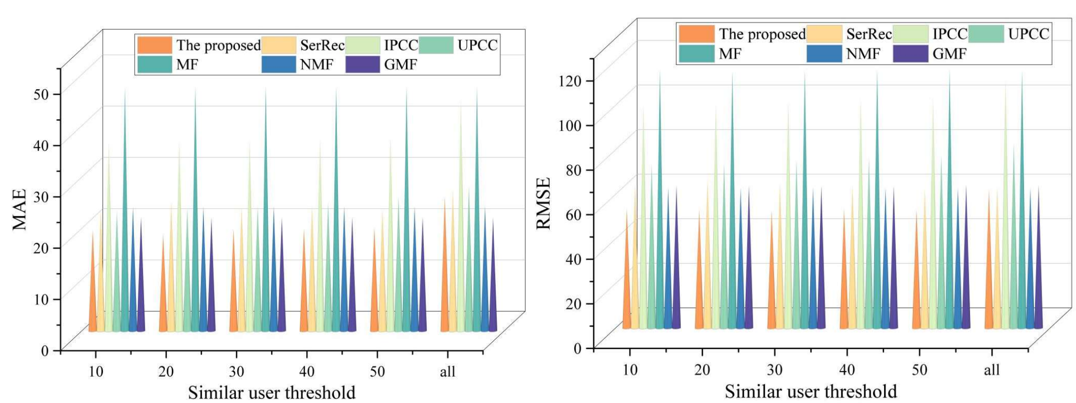


Figure 5: Comparison of MAE and RMSE

The accuracy of the results of the proposed model in this paper for recommendation using different similar user thresholds under different data sparsity is further demonstrated and by comparing it with six other mainstream methods (SerRecd, IPCC, UPCC, MF, NMF, GMF). The effects of taking different Top-K similar users under different data sparsity on the experimental results of the model are shown in Table 2. Specifically, the sparsity Sparsity value of the dataset is set to 10%, 20%, 30%, 40%, 50%, 60%, 70%, and 80%, which means that the above nine proportions of the original data are randomly erased from the original throughput dataset, respectively; and under the datasets with different sparsities, different similar user thresholds are selected {10,20,30,40,50,al1} for recommendation. When the data sparsity increases from 10% to 80%, the model shows a monotonically increasing trend in both MAE and RMSE metrics. At 10% sparsity, the MAE of Top-10 neighbors is 17.583, and the RMSE is 50.183; while when the sparsity is increased to 80%, the MAE of the same number of neighbors increases to 28.973, and the RMSE reaches 73.083.This phenomenon stems from the insufficient interaction information under sparse data, which leads to the difficulty of the model in capturing the effective latent features, and thus exacerbates the prediction error.



(a)MAE (b)RMSE

Figure 6: Comparison of MAE and RMSE values of the seven methods

Table 2: Accuracy Analysis at different densities

Neighbor Number	Sparsity							
	10%		20%		30%		40%	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
10	17.583	50.183	17.694	52.483	19.492	56.382	23.974	62.483
20	18.084	50.974	18.272	51.048	19.375	56.171	23.686	62.194
30	19.463	52.287	19.592	52.492	20.084	57.392	23.221	61.937
40	20.483	54.366	20.681	55.037	21.228	58.229	24.987	62.574
50	22.779	57.972	22.993	58.135	23.482	59.864	25.385	62.755
ALL	26.037	62.431	26.037	62.431	26.037	62.431	26.037	62.431
	50%		60%		70%		80%	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
10	23.636	62.483	26.998	68.972	27.397	70.837	28.973	73.083
20	23.382	62.311	26.783	64.835	26.984	68.378	28.598	72.984
30	23.011	62.024	26.011	63.287	26.573	65.302	27.387	71.033
40	23.408	62.407	26.183	62.874	26.328	63.197	26.971	68.489
50	23.602	62.459	26.565	63.289	26.184	62.994	26.351	65.386
ALL	26.037	62.431	26.037	62.431	26.037	62.431	26.037	62.431

Second, in order to comprehensively evaluate the performance of the model proposed in the paper under different similar user thresholds, the sparsity of the dataset is uniformly set to 20% in this experiment. A series of similar user values are selected for the experiment, including {10,20,30,40,50,ALL}, and the corresponding MAE and RMSE comparison results are shown in Fig. 6(a~b), respectively. As can be seen from the data in Fig. 6, the proposed model in this paper exhibits good MAE and RMSE values under different numbers of Top-K neighbors, and in the MAE dimension, the proposed model reaches the lowest error value when the number of neighbors is 20, which is only 18.891.In the RMSE dimension, the proposed model also maintains the leading edge, and its RMSE value at the threshold 20 is 53.348, which is better than the suboptimal method NMF (RMSE = 62.986) by about 15.3%.

III. C. Application evaluation

The practicality and scientificity of the model need to be tested and improved and enhanced in teaching practice. For this reason, this paper conducted a trial and survey in a small scope, the trial survey object is the students of English majoring in University A, class 3 and 4, grade 2023, a total of 208 people participated in the trial and questionnaire survey, the recovery rate of the questionnaire is 100%, and the validity rate is 96.15%. In this paper, 200 valid questionnaires are analyzed, and the statistical results of the questionnaire data are shown in Table 3 and Table 4.

Table 3: Results of Questionnaire Survey 1 after the trial

Question		Result		
Q1	Do you think of the interface design of this system Concise?	A.Very good	B.General	C.Uncertain
		94%	5%	1%
Q2	Do you think the structural design of this system is reasonable?	A.Very good	B.General	C.Unreasonable
		84.5%	6.5%	9%
Q3	Do you think this system can effectively help you improve your English proficiency?	A.Yes	B.No	C.Uncertain
		85.5%	8%	6.5%
Q4	Do you use this system to train your oral English?	A.Often	B.Very few	C.No
		84%	10.5%	5.5%
Q5	Have you utilized the network resources provided by this system?	A.Often	B.Very few	C.No
		88.5%	11.5%	0%
Q6	Whether the network resources of this system can meet your learning expectations?	A.Yes	B.General	C.Not satisfied
		85%	11%	4%
Q7	Do you think the functions of this system are relatively reasonable?	A.Reasonable	B.General	C.Unreasonable
		88.5%	10%	1.5%
Q8	Do you think this system is beneficial to you?	A.Very useful	B.General	C.No use
		82.5%	13%	4.5%

Table 3 shows the results of questionnaire survey 1, which mainly explores the user's satisfaction with the technical effects of the system. The questionnaire data shows that 94% of the users choose “very good” for the simplicity of the system interface design, 88.5% recognize the rationality of the system's functionality, 88.5% believe that the network resources meet the learning expectations and 82.5% of the users approve of the system's utility, which demonstrates the effectiveness of the model in semantic matching and real-time interaction.

Table 4 shows the results of Questionnaire 2, which mainly explored the users' experience of the system. 87.5% of the students thought that the system could stimulate the interest in learning English cultural communication, and 88.5% recognized the improvement of English cultural communication skills by combining the learning with the system.

Table 4: Results of Questionnaire Survey 2 after the trial

Question		Result		
Q9	Do you think this system plays a significant role in the study of English cultural communication?	A.Very useful	B.General	C.No use
		83%	13.5%	3.5%
Q10	Can this system bring you a pleasant mood and enhance your interest in learning English?	A.Yes	B.No	C.Uncertain
		87.5%	11%	1.5%
Q11	Do you think that the combined use of this system in class learning can effectively improve your English cultural communication skills?	A.Yes	B.General	C.No
		88.5%	10%	1.5%
Q12	Are you adapted to using this system for self-study of English cultural communication?	A.Very adaptable	B.General	C.Uncertain
		81%	19%	0%

IV. Conclusion

In this paper, we design an English cultural communication system based on a dynamic planning model, utilize experiments to verify the effectiveness of the system, and explore user evaluation through a small-scale trial.

The OCDSLP algorithm using semantic similarity yields high overlap between typical tuples and representative tuples labeled by users, with an accuracy of 85.1% and 84.3% on UsedcarDB and IMDB datasets, respectively (the accuracy is calculated by taking the average of 10 test queries), whereas the selection accuracy of the traditional sorting method is only 52.0% and 56.9%.

The test results on the HelloTalk dataset show that the proposed model outperforms the model without label preassignment using the OCDSLP algorithm under 50% data sparsity for each number of neighbors setting. When the data sparsity increases from 10% to 80%, the proposed model shows a monotonically increasing trend in MAE and RMSE metrics. It performs optimally over the control model under different numbers of Top-K neighbors, and in the MAE dimension, the proposed model reaches the lowest error value of only 18.891 when the number of neighbors is 20.In the RMSE dimension, the proposed model RMSE value is 53.348 at the threshold of 20, which is about 15.3% lower than that of the suboptimal method, NMF (RMSE = 62.986).

Questionnaire data showed that 94% of users chose “very good” for the simplicity of the system's interface design, and 88.5% recognized the rationality of the system's functions. 88.5% thought that the online resources met the learning expectations and 82.5% of the users recognized the usefulness of the system, indicating the effectiveness of the model in semantic matching and real-time interaction. 87.5% of the students thought that the system could stimulate the interest in learning English cultural communication and 88.5% recognized the improvement of English cultural communication skills by combining the learning with the system, which indicated that the users had a better overall experience when experimenting with the system.

Funding

Phase results of the 2022 Lyu Liang University Doctoral Research Initiation Project "A Comparative Study of Chinese-ASEAN Folktales" (UP202-0304405); Phase results of the 2024 Shanxi Provincial Philosophy and Social Sciences Planning Project "Research on Translation, Introduction and International Communication of Sanjin Yellow River Culture in the Digital Intelligence Era" (2024QN159).

References

[1] Wei, T. (2024). The role of English in promoting international cultural exchange in the context of big data and deep learning. Journal of Computational Methods in Science and Engineering, 24(1), 369-384.

[2] Soldatenkova, E. N. (2021). Overcoming the Semantic Barrier and Developing the Ability to Use Communicative Tools Part I. Issue. Method. Cultural-Historical Psychology, 17(3).

[3] Malik, S., Qin, H., & Oteir, I. (2021). Perceived Psychological, Linguistic and Socio-Cultural Obstacles: An Investigation of English Communication Apprehension in EFL Learners. International Journal of Instruction, 14(4), 733-752.

[4] Korkut, P., Dolmacı, M., & Karaca, B. (2018). A study on communication breakdowns: Sources of misunderstanding in a cross-cultural setting. Eurasian Journal of Educational Research, 18(78), 139-158.

[5] Kaur, J. (2017). Ambiguity related misunderstanding and clarity enhancing practices in ELF communication. Intercultural Pragmatics, 14(1), 25-47.

[6] Zhang, J. (2024). Implicit Context in Cross-Cultural Communication of Chinese Brands. Cross-Cultural Communication of Chinese Brands, 55-82.

[7] Alian, M., & Awajan, A. (2020). Semantic similarity for english and arabic texts: a review. Journal of Information & Knowledge Management, 19(04), 2050033.

[8] Cao, J., Wang, T., Li, W., & Tao, C. (2019). A method of calculating the semantic similarity between English and Chinese concepts. In Machine Learning and Intelligent Communications: 4th International Conference, MLICOM 2019, Nanjing, China, August 24–25, 2019, Proceedings 4 (pp. 313-324). Springer International Publishing.

[9] Jatnika, D., Bijaksana, M. A., & Suryani, A. A. (2019). Word2vec model analysis for semantic similarities in english words. Procedia Computer Science, 157, 160-167.

[10] Deng, Y., Bai, W., Li, J., Mao, S., & Jiang, Y. (2024). Improving semantic similarity computation via subgraph feature fusion based on semantic awareness. Engineering Applications of Artificial Intelligence, 136, 108947.

[11] Zunlan, X., & Xiaohong, N. (2025). Dynamic BERT-SVM Hybrid Model for Enhanced Semantic Similarity Evaluation in English Teaching Texts. Journal of English Language Teaching and Applied Linguistics, 7(1), 01-11.

[12] Handriani, P., Dinata, A. P., Siregar, F. A., & Siregar, D. Y. (2025). Politeness Strategies in Cross-Cultural Communication: A Pragmatic Approach. Semantik: Jurnal Riset Ilmu Pendidikan, Bahasa dan Budaya, 3(1), 126-132.

[13] Stepanyan, I. K., Dubinina, G. A., Nikolaev, D. A., Kapranova, L. D., & Pashtova, L. G. (2017). Cross-disciplinary case-analyses of investment optimization in a foreign language applying dynamic programming. Revista ESPACIOS, 38(62).

[14] Zhang, D. (2022, December). Literary Language System on Account of Dynamic Programming Algorithm. In 2022 IEEE 2nd International Conference on Mobile Networks and Wireless Communications (ICMNWC) (pp. 1-5). IEEE.