Tourism Destination Recommendation based on Bag of Visual Word Combined with SVM Classification

Xiaohua Liu
School of Economics and Management, Yan'an University, Yan'an, 716000, China
E-mail: ella4312@163.com
*Corresponding author

Keywords: bag of visual word, SVM, tourism recommendation, CNN, DA

Received: June 19, 2024

In recent years, people's living standards have gradually improved. More and more people plan to travel. In response to the low accuracy of user travel destination recommendation, this study proposes a travel destination recommendation method combining bag of visual word with support vector machine. Firstly, the study introduces a convolutional neural network extractor to improve bag of visual word. In the improvement, a convolutional layer is selected and treated as a dense descriptor extractor. This layer is embedded into a visual bag of words model to learn more suitable visual vocabulary. An improved bag of visual word is used to extract feature data from user uploaded online tourism images. A low-level feature set and a high-level semantic feature set of the source domain data are constructed. Subsequently, domain adaptation is introduced to address the distribution differences between the target feature data and the source domain feature data. Finally, the support vector machine is improved to classify the attractions that users are interested in. The similarity calculation is used to achieve tourism destination recommendation. The experimental results showed that the average accuracy of the proposed algorithm was 89.54%, the recall was 60.58%, and the macro F1-score was 89.92%. These values were all better than the comparative algorithms and the bag of visual word before optimization. Overall, the designed tourism destination recommendation algorithm has strong practical applicability. This algorithm provides a strong recommendation strategy for many users to travel and helps them efficiently choose their desired attractions.

Povzetek: Študija uvaja izboljšan sistem priporočanja turističnih destinacij z uporabo "vreč virtualnih svetov" in izboljšan SVM, kar povečuje kvaliteto priporočil in optimizira izbiro destinacij.

1 Introduction

With the accelerated development of tourist attraction construction, people are increasingly accessing tourist attractions from the internet. According to statistics, the number of Chinese tourists in 2023 was 4.891 billion, an increase of 2.361 billion compared to the same period last year and a year-on-year increase of 93.3%. The domestic tourists for urban residents reached 3.758 billion, a year-on-year increase of 94.9% [1-3]. The rural residents traveling domestically reached 1.133 billion, a year-on-year increase of 88.5%. The domestic tourists in the first quarter were 1.216 billion, a year-on-year increase of 46.5%. The domestic tourists in the second quarter were 1.168 billion, a year-on-year increase of 86.9%. The domestic tourists in the third quarter were 1.29 billion, a year-on-year increase of 101.9%. The domestic tourists in the fourth quarter were 1.217 billion, a year-on-year increase of 179.1% [4]. Faced with strong tourism demand, local governments have introduced a series of policies. At the same time, local cultural and tourism bureaus have also carried out a lot of publicity on the internet to promote local tourist attractions. However, people often find it difficult to accurately screen various types of tourism information. The difficulty of obtaining information about target tourist attractions is gradually increasing [5-6]. To address this issue, numerous researchers have designed tourist attraction recommendation systems using recommendation algorithms. The recommendation algorithm uses mathematical methods to infer what users may like. The recommendation algorithm has been applied in many fields, such as movies, books, music, painting, dance, learning, etc. The widespread application of recommendation algorithms is not only beneficial for businesses to operate, but also helps users to obtain target items. The effective application of recommendation algorithms in tourism is currently a key focus of development for many tourisms software. This study uses the Bag of Visual Word (BoVW) model combined with Support Vector Machine (SVM) to recommend tourist attractions. The contribution of the research lies in fully integrating the advantages of the proposed algorithm for image feature extraction and effective processing of unlabeled data. This helps to design a more efficient tourist attraction recommendation system.

The research content mainly includes four parts. Firstly, a review is conducted on the tourism destination

recommendation using BoVW combined with SVM. Secondly, the improved BoVW and SVM are introduced. A tourism destination recommendation model is designed. Then this proposed approach is experimentally validated. At last, these experimental outcomes are summarized and future prospects are proposed.

2 Related works

Faced with an increasing amount of tourism information, it is becoming increasingly difficult for people to collect information on target tourist attractions. Numerous researchers have conducted a series of studies to address the strong demand for tourism and build an efficient tourist attraction recommendation system. These are difficulties to extract typical fault characteristics of mixed signals in various applications. Therefore, Long et al. put forward a diagnosing method using image information and BoVW. The diagnostic accuracy of four fault states reached 99.50% [7]. Saini and Susan put forward a novel combination of visual codebook generation using deep features and nonlinear classifiers to address the classification issues. This new method effectively solved the classification of multi-class imbalanced image datasets [8]. Gangrade et al. proposed an identification system to bridge the gap between deaf mute individuals and society. The system used Microsoft Kinect sensors to segment hand areas from depth images in cluttered environments. Then the obtained depth image was used to achieve supervised machine learning by extracting and training the features of the image. The proposed system generated an average recognition accuracy of 93.26% for the Indian sign language dataset [9]. To enhance engineering cost prediction accuracy, Fan and Sharma proposed a construction cost prediction model based on SVM and least squares SVM. This prediction model's relative error was within 7%, with high predicting accuracy and stability [10]. Shen and Yan proposed a new method using intermediate domain SVM to predict the remaining service life of rolling bearings. This transferring approach aimed to deal with the poor source degrading index that could not be applied. This new approach had higher RUL predicting outcomes than other means, demonstrating the multi-optimization transfer learning's advantages [11]. Cui et al. put forward a novel recommending means using time correlation coefficient and a modified cuckoo search K-means to meet the fast and accurate recommendation requirements of users in the Internet of Things environment. This new model was effective for the Internet of Things scenario [12]. Zhou et al. proposed an integrated CNN-RNN-based modeling and analysis of patient doctor generated data to extract and highlight the combination of semantic and sequential features in patient inquiries. The proposed model for intelligent pre-diagnostic services in online medical environments was effective [13]. Sun et al. proposed a new method called long-term and short-term preference modeling to address the unreliable recommendation results of existing RNN-based methods for the next point of interest recommendation. This newly proposed model produced significant improvements compared to state-of-the-art methods [14]. To address the high learning efficiency requirements of non-sampling strategies, Chen et al. proposed a general framework called ENMF using a simple neural matrix decomposition architecture. The proposed ENMF framework outperformed state-of-the-art methods significantly in Top-K recommendation tasks [15]. Huang et al. proposed a multi-attention-based group recommendation mode to enhance the effectiveness of recommendations. It effectively utilized a deep neural network structure using multi-attention to achieve accurate group recommendations. This new model was significantly superior to advanced methods in dealing with recommending problems [16]. The summary of related works is shown in Table 1.

Table 1: Summary of related works

Author	Accuracy	Recall	F1-score	Author	Accuracy	Recall	F1-score
Z. Long	85.46%	58.49%	80.52%	Z. Cui	83.25%	56.55%	84.44%
M. Saini	84.18%	55.24%	81.43%	X. Zhou	86.64%	55.89%	85.17%
J. Gangrade	86.42%	54.49%	85.49%	K. Sun	84.82%	56.49%	85.67%
M. Fan	87.46%	56.29%	86.17%	C. Chen	87.45%	58.48%	86.18%
F. Shen	85.93%	58.18%	83.47%	Z. Huang	88.56%	57.85%	88.64%

In summary, numerous scholars have conducted extensive research on the application of recommendation algorithms. Based on this, the improved BoVW and SVM are applied to the tourist attraction recommendation model to solve the current problem of inaccurate and inefficient tourism recommendations.

3 BoVW combined with SVM for tourism destination recommendation

The study proposes a tourism destination recommendation system that combines BoVW with SVM classification to address the insufficient accuracy of traditional tourism destination recommendation systems. Firstly, a Convolutional Neural Network (CNN) extractor is introduced to improve BoVW. Feature data extraction is performed on online travel images uploaded by users by improving BoVW. A low-level feature set and a high-level semantic feature set are constructed for Source Domain (SD) data. Subsequently, Domain Adaptation (DA) is introduced to eliminate the target and SD feature data's distribution differences. Finally, the target and related users' similarity is calculated. Personalized tourism destination recommendations are achieved by improving SVM to classify the attractions that users are interested in.

3.1 Image data feature extraction based on improved BoVW

BoVW is mainly based on the original idea of the bag of words model, which is applied in text classification at the beginning. By representing documents as feature vectors to ignore word order, grammar, and syntax, they are only considered as a collection of vocabulary [17-18]. In BoVW, this idea is applied in image processing, describing image content by representing images as a set of visual words. BoVW mainly includes several steps: feature extraction, generating visual dictionaries, mapping to visual dictionaries, training, and classification. Image feature extraction is performed using the Scale Invariant Feature Transform (SIFT) algorithm. BoVW's core is to use local features of an image as visual words and describe the content of the image by counting the frequency of these visual words appearing in the image [19-21]. This method not only preserves the image's local features, but also compresses the image description, making the image representation more concise and easier to process. Figure 1 shows the basic process of BoVW.

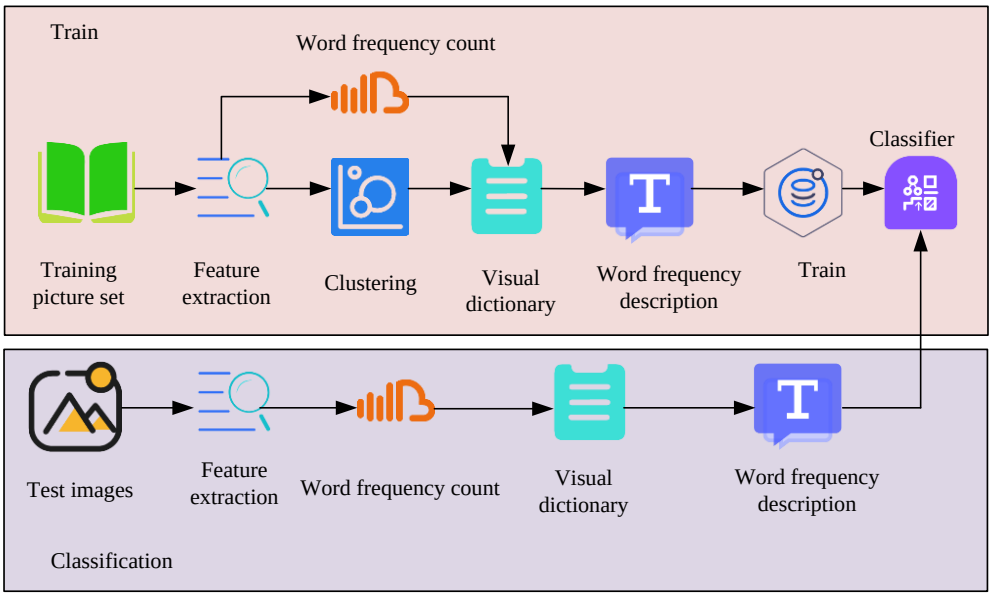


Figure 1: Image classification flowchart of BoVW

In Figure 1, BoVW extracts image features to form a subset of image feature descriptors, which are then clustered to form a visual dictionary. Finally, the image is classified using a classifier. Traditional BoVW faces the low efficiency in extracting features from a mass of data. Therefore, this study introduces deep neural networks to optimize BoVW feature extraction in Figure 2.

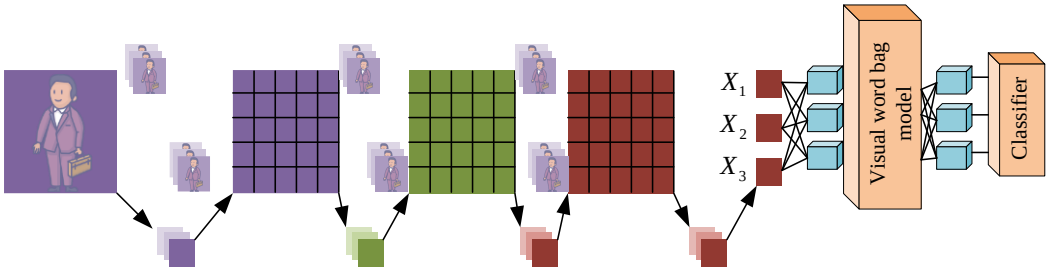


Figure 2: Deep neural network optimization

In Figure 2, a CNN optimized feature extractor is selected to obtain more information from the image. Then a convolutional layer is selected and treated as a dense descriptor extractor, embedded in BoVW learning, which is more suitable for visual vocabulary. Finally, a result is obtained through the trained classifier. The research uses CNN as the ImageNet weight on AlexNet. The first 7 layers of the AlexNet network are not processed. Adaptive metrics are added to the fully connected layer in the last layer, using the maximum mean difference criterion. The first 7 layers of AlexNet consist of 5 convolutional layers (including 3 pooling layers) and 2 fully connected layers. Each convolutional layer contains a convolutional kernel, bias term, ReLU activation function, and local response normalization module. The first, second, and fifth convolutional layers are followed by a max pooling layer. The final output layer is softmax, which converts the network output into probability values for predicting the category of the image. In CNN, the input image is processed layer by layer. Each layer extracts different feature information. These layers can be seen as different filters that recognize specific patterns and shapes in the image, such as edges, corners, and lines [22-23]. As layers gradually increase, CNN is able to extract increasingly complex features, such as textures, shapes, and structures in images. The basic principles of CNN are divided into input layer, convolutional layer, activation function, pooling layer, multi-layer stacking, fully connected layer, and output layer [24]. Convolutional layer is the core of CNN. The input image is convolved with a set of learnable convolution kernels. Convolution operation can be understood as sliding the convolution kernel over the input image and calculating the dot product between the convolution kernel and image's local region to generate a feature map. Each convolution kernel can extract different features, such as edges, textures, etc. In Figure 3, the advantage of convolution operation is that it can reduce network parameters while preserving local spatial relationships, ultimately achieving the goal of enhancing image data feature extraction.

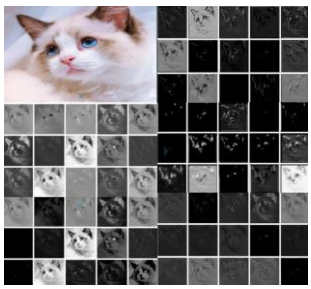


Figure 3: Visualization of CNN image feature extraction

In Figure 3, a picture of a cat is input. The first layer of CNN detects the edges and corners of the cat's body. The second layer extracts the cat's ears shape and the face contour. The third layer further analyzes the texture, eyes, and shape of cat hair. This feature extraction preserves local spatial relationships and reduces network parameters, allowing the extractor to better process image feature information. The image features extracted by CNN at different levels are observed through feature extraction. Moreover, low-level features containing edges and textures are also presented. High-level features include more abstract and semantic information containing eyes, nose, and mouth. By comparing images from different layers, image's features are more abstract and semantically stronger than those of the previous layer. By observing the changes in the features of each layer, edge information becomes more abundant. By using CNN for image recognition, higher recognition accuracy and faster extraction speed can be achieved.

3.2 Distribution difference elimination based on DA technology

The basic feature set of user uploaded network images obtained by optimizing feature extraction of BoVW using

CNN is called SD dataset. Due to the differences between SD dataset and the obtained target dataset, the study uses DA to eliminate these differences. The purpose of using DA is for improving the performance of Target Domain (TD) of unidentified data through a large amount of identified data. Then the distribution difference between the two types of data can be eliminated. Figure 4 shows the DA principle.

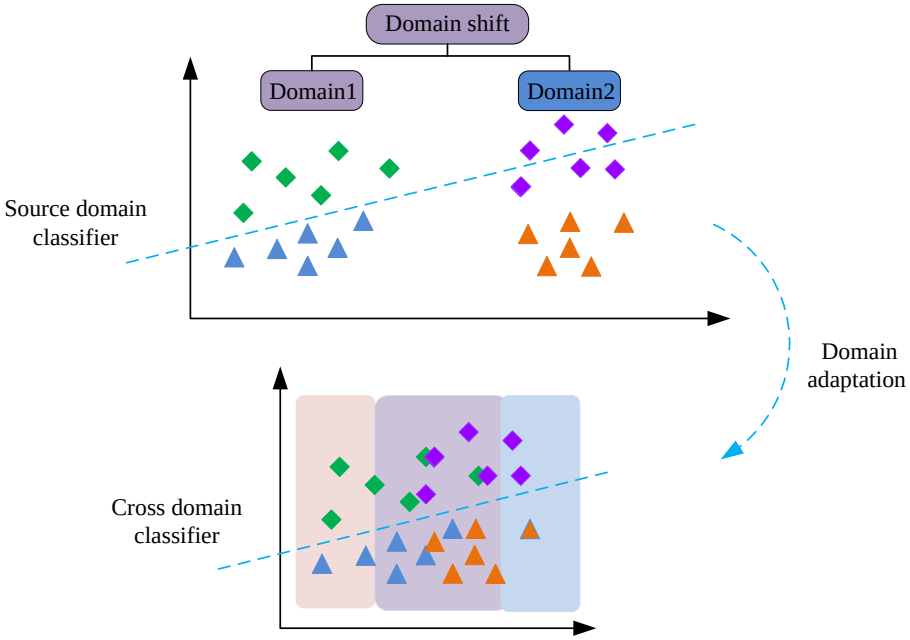


Figure 4: Domain adaptation technology schematic diagram

In Figure 4, DA improves the performance of the model on TD containing insufficiently labeled data by utilizing the knowledge learned from another relevant domain with sufficiently labeled data [25]. The mechanism of DA is to discover the source and the common potential factors of TD. DA adapts them to reduce marginal and conditional mismatches between domains in terms of feature space [26]. DA mainly measures and reduces SD dataset and target dataset's spatial distribution distance by using the maximum mean difference method. This can achieve a reduction in distribution differences between datasets. The mean difference is a loss function defined by equation (1).

$$MMD[F,p,q] := \sup_{f\in F}\left(E_p[f(x)] - E_q[f(y)]\right) \quad (1)$$

In equation (1), MMD refers to the mean difference function. F refers to the domain of functions. x, y correspond to SD data and TD data. p, q correspond to the distributions of x, y. $f(x), f(y)$ correspond to the values of SD after mapping f and TD after mapping f, respectively. E_p refers to expectations. $\sup$ refers to the supremum. Equation (2) can be obtained by transforming equation (1).

$$MMD[F,p,q] := \sup_{\|f\|_{H\le 1}}\begin{pmatrix} E_p\left[\langle f, \varnothing(x)\rangle_H\right] \\ -E_q\left[\langle f, \varnothing(y)\rangle_H\right] \end{pmatrix} \quad (2)$$

In equation (2), $\varnothing$ refers to the mapping function. H stands for the reproducing kernel Hilbert space. Equation (2) utilizes the regenerative property of the reproducing Hilbert space by mapping vectors to the Hilbert space through the function $\varnothing$. The mapping to a high-dimensional transformation is completed by dot product with a given vector within a unit sphere in that space. Equation (3) can be obtained by calculating equation (2).

$$MMD[F,p,q] := \sup_{\|f\|_{H\le 1}}\left[\langle f, \mu_p - \mu_q(x)\rangle_H\right] \quad (3)$$

Equation (3) utilizes the property of inner product. μ_p, μ_q represent the expectations of $\varnothing(x), \varnothing(y)$. The formula for calculating the maximum mean difference can be obtained, represented by equation (4).

$$MMD = \left\| \frac{1}{n}\sum_{i=1}^{n}\varnothing(x_i) - \frac{1}{m}\sum_{j=1}^{m}\varnothing(y_j)\right\|_H \quad (4)$$

In equation (4), n refers to the samples number in a SD. m refers to the number of samples in TD. i, j refer to vectors in the feature space. Due to the different mapping functions in different tasks, a kernel function is introduced for expression, represented by equation (5).

$$MMD^2[F,p,q]=\left\|\frac{1}{n^2}\sum_{i=1}^{n}\sum_{i'=1}^{n}k(x_i,x_i')-\frac{2}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}k(x_i,y_i)+\frac{1}{m^2}\sum_{j=1}^{m}\sum_{j'=1}^{m}k(y_j,y_j')\right\|_H \quad (5)$$

In equation (5), i', j' refer to vectors in different feature spaces. k represents the introduced kernel function. Figure 5 is a framework for eliminating the distribution differences between SD data and target data using DA.

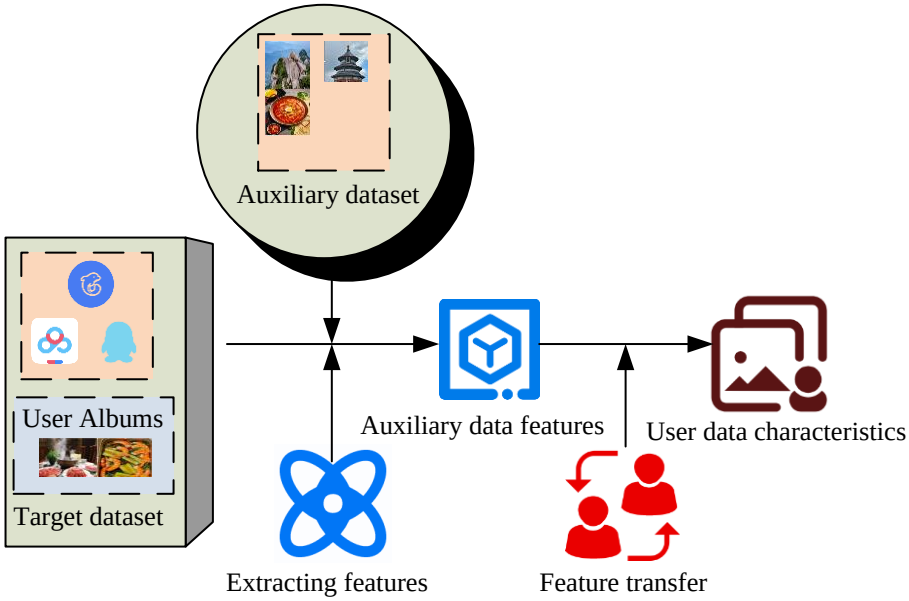


Figure 5: Framework diagram of domain adaptation technology for eliminating data distribution differences

In Figure 5, first, the improved BoVW is utilized to crawl the tourist attraction images uploaded by network users and extract their features as SD dataset. Subsequently, DA is used to eliminate the distribution differences between SD data and target data. The auxiliary dataset features are transferred. User interest recommendations are made based on the target user and other users' similarity. The process of transferring user data features using DA is shown below.

$$S=\{S_i\}_{i=1}^{N} \quad (6)$$

In equation (6), S refers to SD dataset. The i th type of attraction in the category of attractions is represented as S_i. The types of scenic spots are represented by equation (7).

$$S_i=\{I^i\} \quad (7)$$

In equation (7), the i th type of scenic spot image set is represented as I^i. The feature set extracted from the scenic spot image set is represented by equation (8).

$$F^i=\{F_i^i, y_i=i\}_{i=1}^{N} \quad (8)$$

In equation (8), the image feature vector of S_i is represented as F^i. The feature vector of S_i is labeled as i. The TD dataset is represented by equation (9).

$$U=\{u_1,u_2,u_3,\cdots,u_i\} \quad (9)$$

In equation (9), the i th user is represented as u_i. The album data of u_i is represented by equation (10).

$$u_i=I_u^i \quad (10)$$

The set of feature vectors extracted by I_u^i is represented as equation (11).

$$F_u^i=F_{ul}^i\bigcup F_{uu}^i \quad (11)$$

In equation (11), a small amount of labeled data is represented as F_{ul}^i. A large amount of unlabeled data are represented as F_{uu}^i. The maximum mean difference between two data types is represented by equation (12).

$$DIST_k(U,S)=\left\|\frac{1}{M}\sum_{i=1}^{M}\varphi\left(F_u^i\right)-\frac{1}{N}\sum_{i=1}^{N}\varphi\left(F^i\right)\right\|_H \quad (12)$$

In equation (12), $DIST_k$ refers to two data types' distance mapped in the reproducing kernel Hilbert space. M refers to the samples number in the user album dataset. N refers to the sample size in the attraction type dataset.

3.3 Tourism destination recommendation based on SVM

This study improves SVM to classify the scenic spots that users are interested in after using DA to eliminate the distribution differences. Meanwhile, the study calculates

the similarity values between the target and related users to achieve personalized tourism destination recommendations. SVM is a powerful supervised learning means utilized in classification and regression. It can find a hyperplane that maximizes the classifying boundary and the nearest training sample's interval [27]. The points closest to the hyperplane are support vectors. Since these farther points are mainly established through support vectors. Figure 6 is a schematic diagram of a hyperplane.

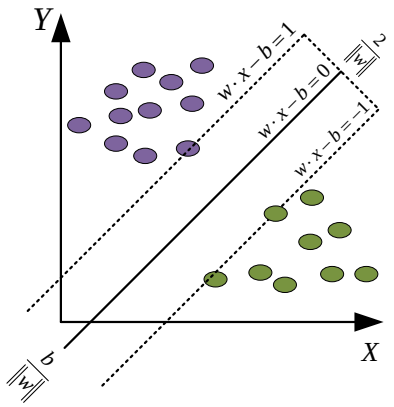


Figure 6: Normal vectors for SVM hyperplane

Figure 6 is a schematic diagram of the normal vector of the SVM hyperplane. w, x, and b correspond to a vector perpendicular to this hyperplane, namely the normal vector, data point, and intercept term. In the binary classification problem shown in Figure 5, SVM is able to find a hyperplane to separate two sets of data and make the data points farthest from the separation interface. SVM can be divided into two situations, namely linearly separable and linearly indivisible. Linear separability refers to using a linear function to separate two samples [28-29]. In practical problems, samples are often not completely linearly separable. Kernel functions need to be introduced for processing. In the low-dimensional space, samples are linearly inseparable. However, in high-dimensional space, samples may be linearly separable. Samples can be divided by increasing the dimension of space [30-31]. Kernel functions can map data to a higher dimensional feature space, making the data linearly separable in this new space. This way, a hyperplane can be constructed in this new space for classification. Commonly used include linear, polynomial, Gaussian, and Laplacian kernel functions. This study employs multi-kernel learning to enhance the effectiveness of kernel functions in practical applications. Multi-kernel learning can make kernel functions more efficient by combining multiple kernel functions. There are various ways to combine multi-kernel learning. Equation (13) is a simple sum.

$$K\left(x_i, x_j\right) = \sum_{n=1}^{O} k_n\left(x_i, x_j\right) \quad (13)$$

In equation (13), O represents the total individual kernel functions. Equation (14) is a weighted sum.

$$K\left(x_i, x_j\right) = \sum_{n=1}^{O} k_n\left(x_i, x_j\right) \quad s.t. \quad d_n \geq 0, \sum_{n=1}^{O} d_n = 1 \quad (14)$$

In equation (14), d_n represents the weight ratio of a single kernel function. Due to the inability of traditional SVM to solve the problem of unlabeled training data, this study introduces Cross Domain Support Vector Machine (CDSVM). CDSVM trains SD data to obtain a hyperplane and support vector, then trains a new hyperplane using TD data, and finally classifies the data in TD using the new hyperplane. CDSVM solves the maximum distance function of the hyperplane using equation (15).

$$y_i\left(w^T \varnothing\left(x_i\right) + b\right) > 1 - \xi_i \quad (15)$$

In equation (15), y_i represents a sample label. w^T means a hyperplane normal vector. x_i means sample data. b means intercept. ξ_i means relaxed variables. CDSVM is used to classify the styles of scenic spots that users are interested in. The categorized scenic spot styles are recommended to the target users through the similarity calculation in Figure 7.

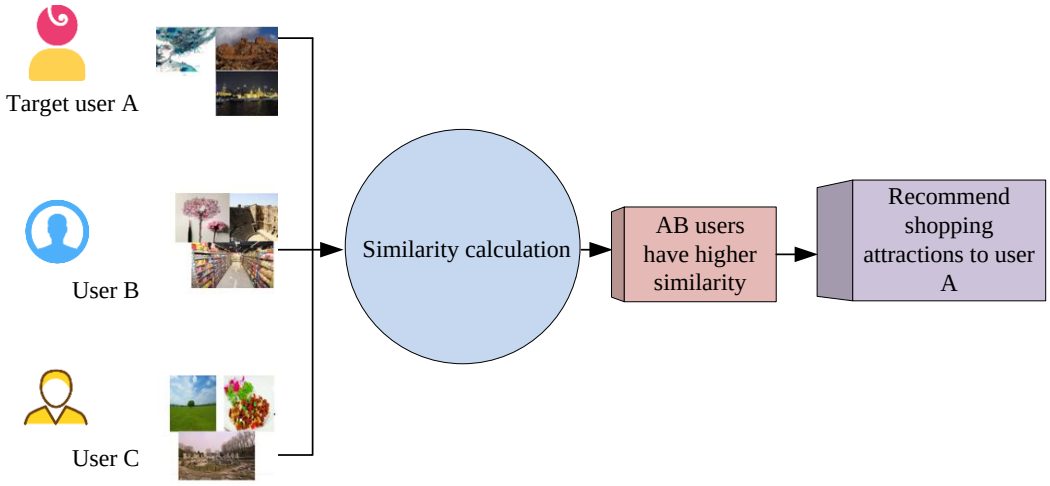


Figure 7: Schematic diagram of user tourist attraction recommendation based on CDSVM

In Figure 7, the user travel destination recommendation model utilizes CDSVM to obtain the classification style of the scenic spots that the user is interested in. Then the model analyzes the styles that different users are interested in and calculates the user relationship with the highest similarity through calculation. Finally, the target user will receive associated attraction style recommendations.

4 Experimental analysis of tourism destination recommendation based on BoVW and SVM

First, the study compared traditional SIFT and CNN optimization algorithms to validate this image data feature extraction method's effectiveness based on improved BoVW. Then the effectiveness of DA was verified. Finally, a feasibility analysis was conducted on the proposed tourist attraction recommendation model.

4.1 Selection of evaluation indicators

The evaluation indicators selected for this experiment include accuracy, recall, F1-score, Root Mean Square Error (RMSE), etc. Accuracy is defined as the percentage of correctly predicted results in the total sample. Accuracy reflects the overall judgment ability of the model on all samples. The meaning of recall is the probability of being predicted as a positive sample among actually positive samples. A high recall means that the model can find as many actually positive samples as possible. The F1-score is calculated by considering both accuracy and recall simultaneously, achieving a balance between both. The balance point on the accuracy and recall curves is the F1-score. A higher F1-score usually means that the model has both high accuracy and high recall when identifying positive samples. RMSE is an indicator used to measure the prediction accuracy of a predictive model on continuous data. RMSE measures the root mean square difference between the predicted and true values, representing the average degree of deviation between the predicted and true values. Accuracy, recall, and F1-score were selected as evaluation indicators to assess the predictive performance of the proposed model in tourism recommendation. Meanwhile, RMSE was introduced to evaluate the accuracy of the model.

4.2 Image data feature extraction analysis based on improved BoVW

The target dataset for this study is the tourism dataset provided by Flickr Image website, which includes a total of 4762 users and 8 cities. The SD dataset is obtained through network crawling. The classification of the TD dataset can be based on the SD data. The experiment adopted the cross-validation method, selecting 80% of the training data for random sampling and the remaining 20% for testing. The experiment used TensorFlow framework as the platform. CNN adopted the AlexNet model pre-trained by ImageNet. Its parameters were set to 105 iterations, momentum factor set to 0.9, attenuation parameter set to 0.006, initial learning rate set to 0.001, and other parameters remain unchanged. The SVM classifier was implemented through Scikit learn, with a penalty coefficient set to 0.005 and a Gaussian kernel function selected. The TD data used in the experiment are shown in Table 2.

Table 2: Target domain dataset information

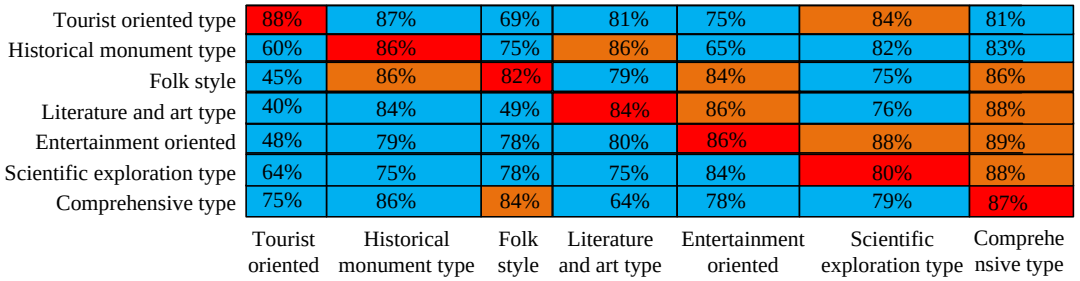
City	Quantity
Berlin	16864
Moscow	14536
London	20548
Los Angeles	43546
Rome	14259
New York	35872
Barcelona	25684
Chicago	19564

Table 3 presents the experimental parameters.

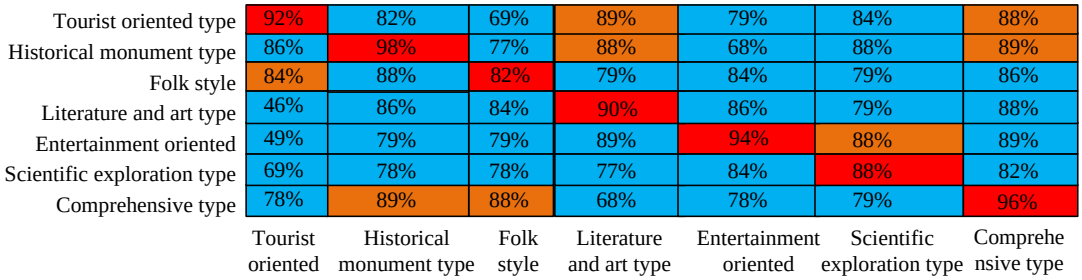
Table 3: Experimental parameters

Name	parameters
Processor	Intel(R) Xeon(R) Cpu E5-2650v2
GPU	NVIDIA Geforce GTX 1650Ti
Memory	16GB DDR4
Hard disk	256G SSD
Operating system	Windows 10
Programming language	Python 3.9
Computer Vision Library	OpenCV
Machine learning library	SciPy
Deep learning framework	Tensorflow

The experiment selected traditional SIFI and CNN-based feature extractor algorithms for comparative experiments to verify the image data feature extraction method based on improved BoVW. A total of 3546 photos from 7 different types of tourist attractions were selected for feature extraction in Figure 8.



(a) SIFI feature recognition results



(b) Improved CNN feature rxtractor recognition results

Figure 8: Comparison of feature extraction's accuracy for seven images

In Figure 8 (a), the CNN improved feature extractor had a recognition accuracy of 88% for tourist-oriented type, 86% for historical monument type, 82% for folk type, 84% for literature and art type, 86% for entertainment-oriented type, 80% for scientific exploration type, and 90% for comprehensive type. In Figure 8 (b), the improved CNN feature extractor had a recognition accuracy of 92% for tourist-oriented type,

98% for historical monument type, 82% for folk type, 90% for literature and art type, 94% for entertainment-oriented type, 88% for scientific exploration type, and 96% for comprehensive type. The improved CNN feature extractor had a higher accuracy in extracting image features. Three classic CNNs and an improved feature extractor were selected for the experiment to predict and compare image features in Figure 9.

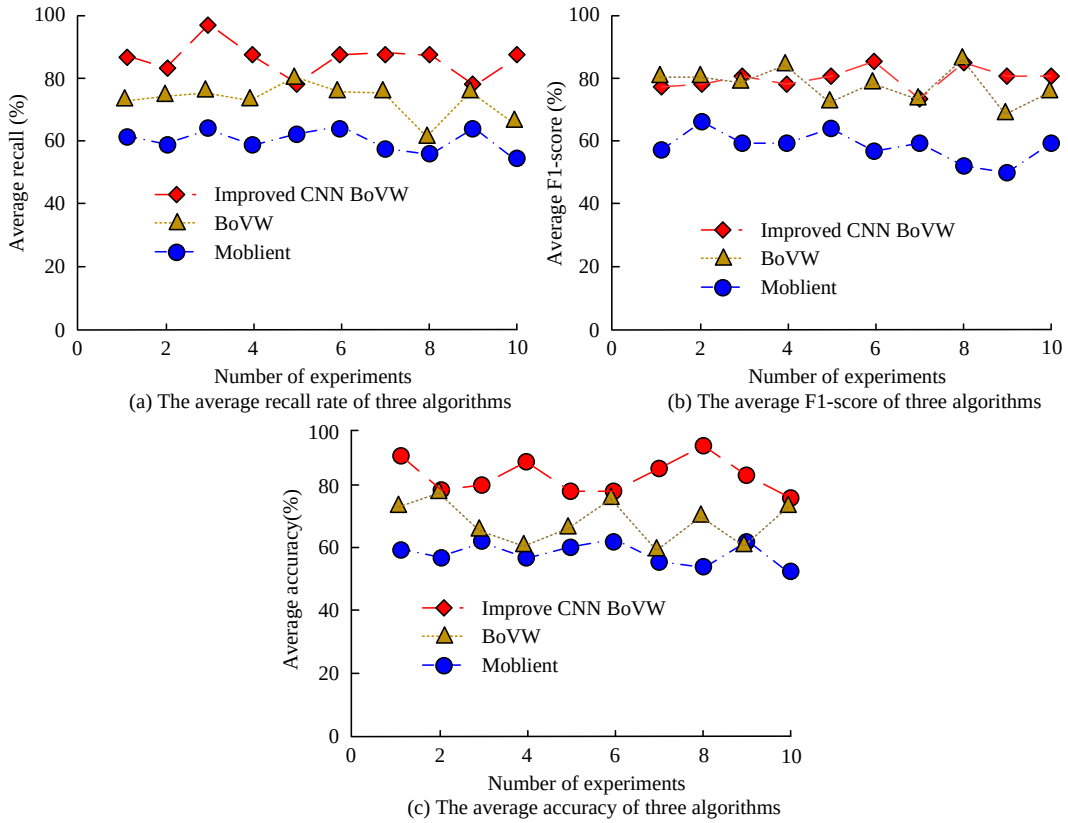


Figure 9: Improved CNN feature extractor for feature prediction comparison chart

In Figure 9 (a), the average recall of the proposed algorithm was 89.64%, BoVW of 78.84%, and Moblient of 60.58%. In Figure 9 (b), the macro F1-score of the proposed algorithm was 89.92%, BoVW of 87.64%, and Moblient of 60.89%. In Figure 9 (c), the proposed algorithm had an accuracy of 89.54%, BoVW of 78.81%, and Moblient of 61.12%. The proposed algorithm had the highest accuracy in feature extraction and stronger performance.

4.3 Analysis of data difference elimination methods based on DA

DA can map data from source and TDs with different distributions to a feature space, making these two sides' distance close in that space. This objective function trained on SD can be transferred to TD to improve accuracy in TD. Figure 10 shows a comparison of data before and after DA.

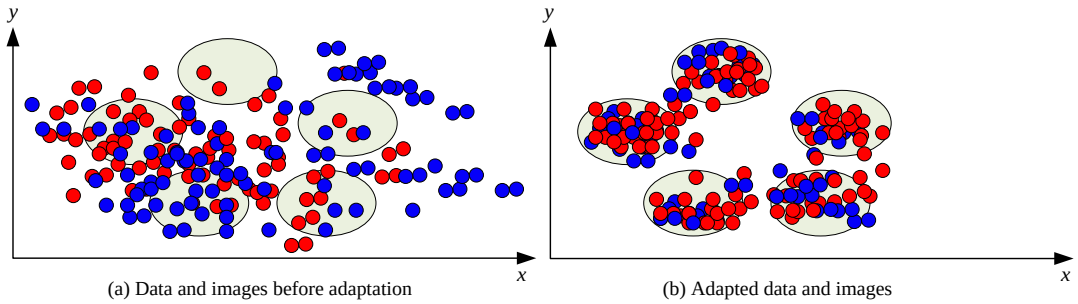
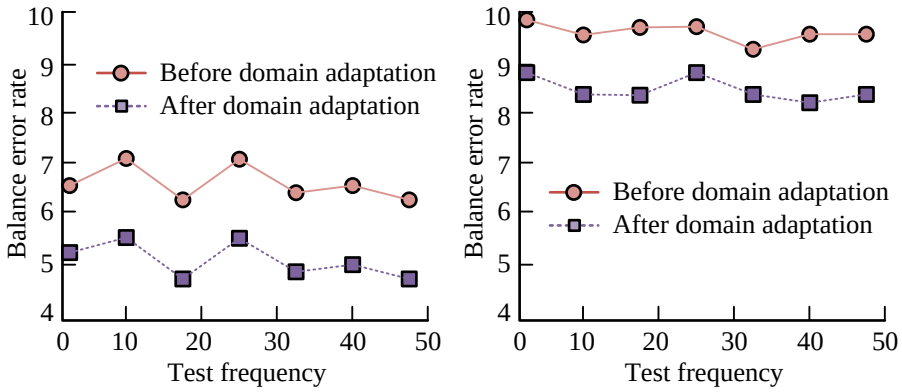


Figure 10: Domain adaptation renderings

In Figure 10 (a), the distributions of SD data and TD data had a significant difference in the feature space before DA. In Figure 10 (b), after DA, the distribution difference between SD and TD data in the feature space was reduced. The distance in space was close. DA effectively reduced the distribution difference between SD and TD data in the feature space. To validate this proposed DA's effectiveness, this study used the GTAVS dataset for separate detection training, without conducting DA training. Then, the GTAVS dataset was used as SD. The SBU dataset was used as TD. A DA module was added to train DA. After DA training was over, the GTAVS and SBU test sets were used to test the model separately in Figure 11.



(a) Test results on the GTAVS test set (b) Test results on the SBU test set

Figure 11: Graph of domain adaptation monitoring test results

In Figure 11 (a), after DA training, the model's balance error on the GTAVS test set decreased from 6.46 to 5.52. The detection accuracy achieved a significant improvement. In Figure 11 (b), the model accuracy on the SBU test set increased, and the equilibrium error rate decreased from 10.21 to 9.14. DA training effectively improved the model's detection accuracy in TD. DA had a good effect in solving the distribution difference between SD and TD data. To verify the practical application feasibility of DA in user travel recommendations, an example analysis was conducted in Figure 12.

Figure 12: Example diagram of user travel recommendation based on domain adaptation method

In Figure 12, the user album was classified based on the assistance of SD data. The model obtained the user's preferred style of tourist attractions and determined that the user's preferred tourist attractions were cultural relics and fashion shopping types. Finally, the model recommended tourist attractions such as Terra Cotta Warriors and shopping centers in first tier cities according to users' preferences. DA effectively eliminated the distribution difference between SD and TD data, thereby providing personalized recommendations to users based on a large amount of labeled network data.

4.4 Analysis of tourism destination recommendation based on CDSVM

The experiment selected the tourism dataset provided by Flickr image website as TD data to verify the classification performance of CDSVM. Network crawled images were used as SD data to classify TD data in Table 3.

Table 3: CDSVM classification results

Algorithm	Berlin	Moscow	London	Los Angeles	Rome	New York	Barcelona	Chicago
SVM	43.6%	42.5%	44.9%	48.6%	47.5%	51.8%	46.1%	55.4%
DA	61.8%	67.4%	62.4%	62.4%	61.3%	58.9%	62.2%	67.9%
CDSVM	70.6%	75.6%	77.6%	69.9%	71.4%	77.8%	72.2%	79.5%

In Table 3, the accuracy of SVM classification for attractions such as Berlin, Moscow, London, Los Angeles, Rome, New York, Barcelona, and Chicago were 43.6%, 42.5%, 44.9%, 48.6%, 47.5%, 51.8%, 46.1%, and 55.4%, respectively. The accuracy of DA's classification of attractions such as Berlin, Moscow, London, Los Angeles, Rome, New York, Barcelona, and Chicago were 61.8%, 67.4%, 62.4%, 61.3%, 58.9%, 62.2%, and 67.9%, respectively. The accuracy of CDSVM in classifying attractions such as Berlin, Moscow, London, Los Angeles, Rome, New York, Barcelona, and Chicago was 70.6%, 75.6%, 77.6%, 69.9%, 71.4%, 77.8%, 72.2%, and 79.5%, respectively. CDSVM performed better and performs better in classifying scenic spots. The experiment selected RMSE as the evaluation index and compared different methods for recommendation. The comparative algorithms used were Factorization Machine (FM) and Content-Based Recommendation (CB). Figure 13 shows the experimental results.

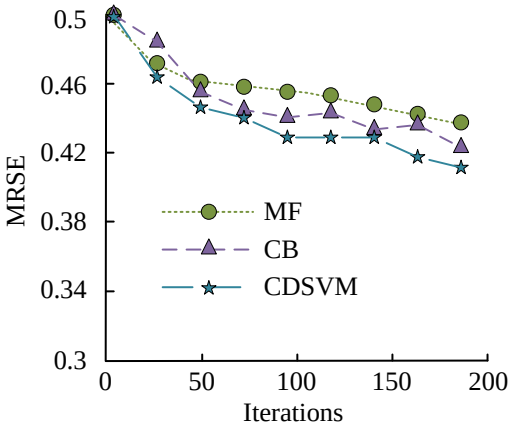
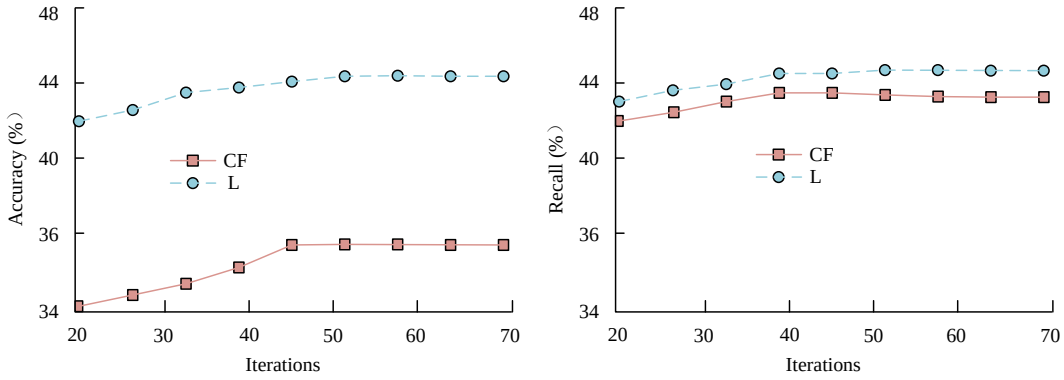


Figure 13: Comparison chart of the performance test of CDSVM

In Figure 13, the RMSE of the comparison algorithm FM gradually decreased with the increasing iterations. The RMSE of 200 iterations was 0.4468. After 200 iterations, the RMSE of CB was 0.4186. The proposed algorithm had an RMSE of 0.3877 after 200 iterations. The RMSE of CDSVM was smaller compared to the other two algorithms, indicating that CDSVM was more stable and had stronger performance. To verify the effectiveness of tourism destination recommendation based on CDSVM, Collaborative Filtering (CF) was used to compare with the proposed algorithm in Figure 14.



(a) Comparison of accuracy between two algorithms (b) Comparison of recall rates between two algorithms

Figure 14: Comparison chart of recommended results

In Figure 14 (a), L represents the proposed algorithm. The nearest neighbors were 20, 30, 40, 50, 60, and 70. The accuracy of CF started to increase from 34.5% and reached 36.6% when the nearest neighbor users were 50, and then stabilized. The accuracy of the proposed recommendation algorithm was 42.6% when the nearest neighbor users were 20 and 44.9% when the nearest neighbor users were 50. In Figure 14 (b), the recall of CF was 42.1% when the nearest neighbor users were 20 and 43.6% when the nearest neighbor users reached 50. The recall of the proposed recommendation algorithm started to increase from 43.8% and stabilized at around 44.8% when the nearest neighbor users reached 50. In summary, the proposed recommendation algorithm was superior to CF and had better recommendation performance.

5 Conclusion

Currently, there is a problem of low accuracy in recommending tourist attractions. To achieve efficient and accurate tourism destination recommendation, the study optimized the image feature extractor. Subsequently, the distribution differences between different data were eliminated. Finally, a tourism destination recommendation model based on BoVW and SVM classification was constructed. In the results, the improved CNN feature extractor had a recognition accuracy of 92% for tourist-oriented type, 98% for historical monument type, 82% for folk type, 90% for literature and art type, 94% for entertainment-oriented type, 88% for scientific exploration type, and 96% for comprehensive type. The improved CNN feature extractor had an average recall of 89.64%, a macro F1-score of 89.92%, and an accuracy of 89.54%. The SD and target data had a significant distribution difference in the feature space before DA. After the DA method, the distribution difference in the feature space was reduced, and the distance was close. The accuracy of CDSVM in classifying attractions such as Berlin, Moscow, London, Los Angeles, Rome, New York, Barcelona, and Chicago was 70.6%, 75.6%, 77.6%, 69.9%, 71.4%, 77.8%, 72.2%, and 79.5%, respectively. Meanwhile, after 200 iterations, the RMSE of CDSVM was 0.3877. The improved CNN feature extractor had higher accuracy and stronger performance in image feature extraction. DA effectively reduced SD data and target data's distribution differences in the feature space. CDSVM performed better in classifying scenic spots. The proposed recommendation model was highly feasible. However, the training data for the tourism destination recommendation model based on BoVW and SVM designed in the study come from websites and networks. Subsequent research can conduct training with different data sources to further improve the model's generalization.

6 Discussion

In the comparative experiment of feature extraction for different algorithms, the improved CNN algorithm adds adaptive metrics in the last layer on the basis of AlexNet network. The improved CNN is embedded into BoVW to enhance its feature extraction ability. Compared with traditional SIFI feature extraction, the improved CNN feature extractor had an accuracy of 92% for tourist-oriented type, 98% for historical monument type, 82% for folk type, 90% for literature and art type, 94% for entertainment-oriented type, 88% for scientific exploration type, and 96% for comprehensive type. This indicated that the feature extraction performance of this method was better. Long Z et al. proposed a motor fault diagnosis method based on image visual information and bag of words model for fault diagnosis. This method also utilized BoVW for image information extraction, with an accuracy rate of around 90% for image recognition. This indicated that the improved BoVW was more effective in feature extraction. The proposed CDSVM algorithm had a classification accuracy ranging from 70% to 80% in classifying different cities. Shen F et al. proposed a new rolling bearing RUL prediction transfer model based on CDSVM to address the problem of predicting the remaining useful life of bearing structures. The accuracy of model classification is around 70%. This demonstrated the superiority of the proposed CDSVM in scenic spot classification. However, the proposed recommended

model's evaluation indicators are only conventional evaluation indicators. More indicators can be considered for multi-angle evaluation in the future.

Funding Statement

The research is supported by: First-class postgraduate courses of Yan'an University (YJG202302).

References

[1] K. S. Arun, V. K. Govindan, and S. D. M. Kumar, "Enhanced bag of visual words representations for content-based image retrieval: a comparative study," Artificial Intelligence Review, vol. 53, no. 3, pp. 1615-1653, 2020. https://doi.org/10.1007/s10462-019-09715-6.

[2] Z. N. Sultani, and B. N. Dhannoon, "Modified Bag of Visual Words Model for Image Classification," Al-Nahrain Journal of Science, vol. 24, no. 2, pp. 78-86, June, 2021. https://doi.org/10.22401/ANJS.24.2.11.

[3] M. F. Aslan, A. Durdu, and K. Sabanci, "Human action recognition with bag of visual words using different machine learning methods and hyperparameter optimization," Neural Computing and Applications, vol. 32, no. 12, pp. 8585-8597, 2020. https://doi.org/10.1007/s00521-019-04365-9.

[4] K. Amiri, M. Farah, and U. M. Leloglu, "BoVSG: bag of visual SubGraphs for remote sensing scene classification," International Journal of Remote Sensing, vol. 41, no. 5, pp. 1986-2003, 2020. https://doi.org/10.1080/01431161.2019.1681602.

[5] M. A. Chandra, and S. S. Bedi, "Survey on SVM and their application in image classification," International Journal of Information Technology, vol. 13, no. 5, pp. 1-11, 2021. https://doi.org/10.1007/s41870-017-0080-1.

[6] D. M. Abdullah, and A. M. Abdulazeez, "Machine learning applications based on SVM classification a review," Qubahan Academic Journal, vol. 1, no. 2, pp. 81-90, 2021. https://doi.org/10.48161/qaj.v1n2a50.

[7] Z. Long, X. Zhang, D. Song, Y. Tang, S. Huang, and W. Liang, "Motor fault diagnosis using image visual information and bag of words model," IEEE Sensors Journal, vol. 21, no. 19, pp. 21798-21807, 2021. https://doi.org/110.1109/JSEN.2021.3102019.

[8] M. Saini, and S. Susan, "Bag-of-Visual-Words codebook generation using deep features for effective classification of imbalanced multi-class image datasets," Multimedia Tools and Applications, vol. 80, no. 14, pp. 20821-20847, 2021. https://doi.org/10.1007/s11042-021-10612-w.

[9] J. Gangrade, J. Bharti, and A. Mulye, "Recognition of Indian sign language using ORB with bag of visual words by Kinect sensor," IETE Journal of Research, vol. 68, no. 4, pp. 2953-2967, 2022. https://doi.org/10.1080/03772063.2020.1739569.

[10] M. Fan, and A. Sharma, "Design and implementation of construction cost prediction model based on svm and lssvm in industries 4.0," International Journal of Intelligent Computing and Cybernetics, vol. 14, no. 2, pp. 145-157, 2021. https://doi.org/10.1108/IJICC-10-2020-0142.

[11] F. Shen, and R. Yan, "A new intermediate-domain SVM-based transfer model for rolling bearing RUL prediction," IEEE/ASME Transactions on Mechatronic, vol. 27, no. 3, pp. 1357-1369, 2021. https://doi.org/10.1109/TMECH.2021.3094986.

[12] Z. Cui, X. Xu, X. U. E. Fei, X. Cai, Y. Cao, W. Zhang, and J. Chen, "Personalized recommendation system based on collaborative filtering for IoT scenarios," IEEE Transactions on Services Computing, vol. 13, no. 4, pp. 685-695, 2020. https://doi.org/10.1109/TSC.2020.2964552.

[13] X. Zhou, Y. Li, and W. Liang, "CNN-RNN based intelligent recommendation for online medical pre-diagnosis support," IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 18, no. 3, pp. 912-921, May, 2020, https://doi.org/10.1109/TCBB.2020.2994780.

[14] K. Sun, T. Qian, T. Chen, Y. Liang, Q. V. H. Nguyen, and H. Yin, "Where to go next: Modeling long-and short-term user preferences for point-of-interest recommendation," In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 1, pp. 214-221, 2020, https://doi.org/10.1609/aaai.v34i01.5353.

[15] C. Chen, M. Zhang, Y. Zhang, and S. Ma, "Efficient neural matrix factorization without sampling for recommendation," ACMTransactions on Information Systems, vol. 38, no. 2, pp. 1-28, 2020. https://doi.org/10.1145/3402521.

[16] Z. Huang, X. Xu, H. Zhu, and M. Zhou, "An efficient group recommendation model with multiattention-based neural networks," IEEE Transactions on Neural Networks and Learning Systems, vol. 31, no. 11, pp. 4461-4474, 2020. https://doi.org/10.1109/TNNLS.2019.2955567.

[17] A. Kurani, P. Doshi, A. Vakharia, and M. Shah, "A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting," Annals of Data Science, vol. 10, no. 1, pp. 183-208, 2023. https://doi.org/10.1007/s40745-021-00344-x.

[18] R. Sharma, and A. Sungheetha, "An efficient dimension reduction-based fusion of CNN and SVM model for detection of abnormal incident in video surveillance," Journal of Soft Computing Paradigm, vol. 3, no. 2, pp. 55-69, 2021. https://doi.org/10.36548/jscp.2021.2.001.

[19] M. Norouzi Shad, M. Maadani, and M. Nesari Moghadam, "GAPSO-SVM: an IDSS-based energy-aware clustering routing algorithm for IoT perception layer," Wireless Personal

Communications, vol. 126, no. 3, pp. 2249-2268, 2022. https://doi.org/10.1007/s11277-021-09051-5.

[20] M. Reig, A. Forner, J. Rimola, J. Ferrer-Fàbrega, M. Burrel, Á. Garcia-Criado, and J. Bruix, "BCLC strategy for prognosis prediction and treatment recommendation: The 2022 update," Journal of Hepatology, vol. 76, no. 3, pp. 681-693, 2022. https://doi.org/10.1016/j.jhep.2021.11.018.

[21] Y. Zhang, and X. Chen, "Explainable recommendation: A survey and new perspectives," Foundations and Trends in Information Retrieval, vol. 14, no. 1, pp. 1-101, 2020. https://doi.org/10.1561/1500000066.

[22] L. Wu, J. Li, P. Sun, R. Hong, Y. Ge, and M. Wang, "A neural influence and interest diffusion network for social recommendation," IEEE Transactions on Knowledge and Data Engineering, vol. 34, no. 10, pp. 4753-4766, 2020. https://doi.org/10.1109/TKDE.2020.3048414.

[23] S. S. Khanal, P. W. C. Prasad, A. Alsadoon, and A. Maag, "A systematic review: machine learning based recommendation systems for e-learning," Education and Information Technologies, vol. 25, no. 4, pp. 2635-2664, 2020. https://doi.org/10.1007/s10639-019-10063-9.

[24] A. H. Krist, K. W. Davidson, C. M. Mangione, M. J. Barry, M. Cabana, and A. B. Caughey, "US Screening for lung cancer: US Preventive Services Task Force recommendation statement," JAMA, vol. 325, no. 10, pp. 962-970, 2021. https://doi.org/10.1001/jama.2021.1117.

[25] K. W. Davidson, M. J. Barry, C. M. Mangione, M. Cabana, A. B. Caughey, and E. M. Davis, "Screening for colorectal cancer: US Preventive Services Task Force recommendation statement," JAMA, vol. 325, no. 19, pp. 1965-1977, 2021. https://doi.org/10.1001/jama.2021.6238.

[26] D. Alita, A. D. Putra, and D. Darwis, "Analysis of classic assumption test and multiple linear regression coefficient test for employee structural office recommendation," IJCCS, vol. 15, no. 3, pp. 295-306, 2021. https://doi.org/10.22146/ijccs.65586.

[27] L. Qi, W. Lin, X. Zhang, W. Dou, X. Xu, and J. Chen, "A correlation graph based approach for personalized and compatible web apis recommendation in mobile app development," IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 6, pp. 5444-5457, April, 2022, DOI: https://doi.org/10.1109/TKDE.2022.3168611.

[28] K. W. Davidson, M. J. Barry, and C. M. Mangione, "Aspirin use to prevent cardiovascular disease: US preventive services task force recommendation statement," JAMA, vol. 327, no. 16, pp. 1577-1584, 2022. https://doi.org/10.1001/jama.2022.4983.

[29] X. Xia, H. Yin, J. Yu, Q. Wang, L. Cui, and X. Zhang, "Self-supervised hypergraph convolutional networks for session-based recommendation," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, no. 5, pp. 4503-4511, 2021. https://doi.org/10.1609/aaai.v35i5.16578.

[30] C. Ma, L. Ma, Y. Zhang, J. Sun, X. Liu, and M. Coates, "Memory augmented graph neural networks for sequential recommendation," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 4, pp. 5045-5052, 2020. https://doi.org/10.1609/aaai.v34i04.5945.

[31] P. Preethi and H. R. Mamatha, "Region-based convolutional neural network for segmenting text in epigraphical images," Artificial Intelligence and Applications, vol. 1, no. 2, pp. 119-127, 2023. https://doi.org/10.47852/bonviewAIA2202293.