Construction and Analysis of Surrounding Travel Demanding Graph Based on Dual Contrastive Learning Text Classification and Graph Neural Network

Guoping Lai, Zhiheng Chi, Fan Pan, Zhihao Xu and Hao Hu

Information Engineering University, Zhengzhou 450001, China

Abstract

Understanding the main information about the current situation of the tourism market has become an urgent need and new trends in the development of the tourism market. In this paper, we use natural language processing technology to analyze the development of tourism around Maoming City, Guangdong Province during the COVID-19 epidemic by means of data mining methods to build a local tourism graph, refine and design models and methods such as RoBERTa-BiGRU-Attention fusion model, dual contrastive learning, BERT-BiLSTM-CRF named entity identification technique, improved Apriori algorithm, GNNLP model based on conventional models and proved the rationality and efficiency of the improved model by comparative test, provide oriented suggestions to help government departments promote tourism and tourism enterprises product supply, optimize resource allocation and explore the market constantly during the epidemic period after scientific analysis and summary.

Keywords

RoBERTa-BiGRU-Attention fusion model, Dual Contrastive Learning, BERT-BiLSTM-CRF, sentiment analysis, the improved Wilson interval method, improved Apriori, GNNLP

1. Introduction

In the circumstance of the regular prevention and control of the COVID-19 epidemic in recent years, there has been a clear shift in the way that tourists consume tourism in China. Nowadays tourists are more likely to choose short distance travel, the local surrounding travel size skyrocketed ushered in the wind. Under such changes, accurate and rapid understanding the preferences and consumer psychology of tourists has a long-term and positive effect on promoting tourism enterprises product supply, optimizing resource allocation and exploring the market constantly.

With the promotion of "Internet+Tourism" services and the boom of self-media, the main source of information in understanding the current situation of tourism market is Online Travel Agency and User Generated Content data, and using Natural Language Processing technology to analyze tourism text has gradually become a trend. Tourism enterprises and tourism administrators need to use NLP technology to discover relevant tourism elements from relevant tourism texts and tourism product reviews, at the same time digging the correlations between elements and implied high-level concepts, thus predicting and mastering consumer psychology as to make better tourism resource allocation.

Facing the above market demand, Zhang Ju [1] et al. proposed a sentiment classification method by fusing Text-Rank and conducted experiments by using deep learning models such as RNN, LSTM, Text-CNN, BERT; Cui Li Ping [2] et al. proposed a directed graph neural network (L-CGNN) model fused with lexical information for named entity identification in the tourism field to extract tourism entities, Zhang Nuo [3] analyzed the tourism text by constructing knowledge graphs.

© 2022 Copyright for this paper by its authors.

ICBASE2022@3rd International Conference on Big Data & Artificial Intelligence & Software Engineering, October 21-23, 2022, Guangzhou, China

EMAIL: 763170952@qq.com* (Guoping Lai*); 1789222652@qq.com (Zhiheng Chi); panfan2022@163.com (Fan Pan); 793485830@qq.com (Zhihao Xu); 2744190810@qq.com (Hao Hu)

ORCID: 0000-0002-8886-1205 (Guoping Lai);0000-0003-1227-8598 (Zhiheng Chi);0000-0002-2492-8976 (Fan Pan); 0000-0002-4487-4717 (Zhihao Xu);0000-0003-4888-6368 (Hao Hu)

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

However, the current analysis of tourism text mainly consists of single task and does not make full use of text data for comprehensive analysis. Therefore, establishing a tourism demand analysis system based on natural language processing technology has become an urgent need and a new trend in the development of tourism market.

This paper examines the following three questions by analysing the demand for peripheral travel in the circumstance with the normalized prevention and control of the COVID-19 epidemic.

- 1. Identify and classify the huge amount of travel-related WeChat articles pushed online.
- 2. Analyse the popularity of numerous tourism products quantitatively and rank them according to their popularity.
- 3. Construct local tourism graph to mine and analyze implied relationships among tourism products.

2. Data Processing

We collected 3385 online travel-related articles by web crawlers from travel-related texts on Sohu News, Tencent News, China Travel Network etc. The numerous tourism product data were obtained from the data files extracted from major tourism websites.

In 2018 Google team released a pre-training model in natural language processing, BERT [4]. It uses large-scale unlabeled corpus training to obtain textual expressions containing rich meanings, which pioneered the pre-training model. This paper uses an improved fusion model of BERT for text classification. However, the input length of BERT is limited to a maximum of 512 characters [4], which also needs to include two flag bits [CLS] and [SEP]; on the other hand, each character may also be divided into several parts after Tokenizer, so the actual input sentence length may be less than 512 [5]. Meanwhile the length of tourism text is generally quite long, if directly truncate the text that exceed the maximum length, some effective information will be lost, which is detrimental to the classification task, so we need to extract the text summary to solve this problem. This paper tried two approaches to extract text summaries: the unsupervised algorithm Text Rank [6] based on graph ranking and the BiGRU [7] model with bidirectional recognition of text.

Rouge is a set of metrics evolved from the recall rate, and its main idea is to compare the algorithmgenerated summaries with the manually generated standard summaries and evaluate the quality of the summaries by measuring their overlapping degree in N-gram, word sequences and word pairs. Rouge contains Rouge-N, Rouge-L, Rouge-W, and Rouge-S 4 indicators. The comparison of the two Rouge metrics is shown in Table 1 below, which reveals that BiGRU has better results for generative summarization of text and generative summaries work better, have the advantage of synthesizing fulltext information and incorporating external perceptions compared to extractive summaries.

Table 1

Comparison of recall rates of two summary algorithms

	Rouge-1	Rouge-2	Rouge-L
Text Rank	0.243	0.353	0.462
BiGRU	0.509	0.419	0.613

Table 2

Example of data text summary results comparison

ID	Text Content	Text Rank Summary	BiGRU Summary
		Results	Results
001	Spring Festival ticket booking has tips	Spring Festival ticket	Spring Festival is still
	Spring Festival is still more than a month	booking has tips	more than a month
	away, online travel website Spring	Spring Festival is still	away, travel tickets
	Festival ticket sales hot, some routes	more than a month	are hard to find
	and even a ticket is difficult to find	away or can still buy	Wish you all friends
	Miss Xie 13902544039 (608 words)	low discount tickets.	can book the Chinese

		Welcome to attention.(321 words)	New Year air tickets. (289 words)
002	Winter travel know how much 9.6 million square kilometers of the motherland, the four seasons have a unique beauty of winter travel also has a special flavor Miss Xie 13902544039 (1190 words)	Winter travel also has a special flavor. But the harsh winter climate discourages many people physical strength will decline. (323 words)	Although winter travel has its own flavor, the cold weather is a deterrent(313 words)

In the original comment text, there are comment texts with the same content but different IDs, and the duplicate comments are sorted and filtered by time to keep the earliest comments posted. As the travel guide text is unstructured text, it is not uniform in structure with hotel, restaurant and scenic spot comments data. To accurately extract tourism products from unstructured travelogue text data, named entity identification is required, and each sentence of a travelogue guide may contain entities. If using the text summarization algorithm to compress the travel tips, a large number of valid entities may be lost. Therefore, it is necessary to divide each travelogue guide into sentences.

3. Model building and analysis

3.1.Tourism text classification

3.1.1.Text classification based on RoBERTa-BiGRU-Attention fusion model

Currently, updates on the way machine learning classifies and extracts information from text are changing rapidly, The first RNN can adequately learn the text context information, but it is likely to have the problem of gradient dispersion, which is not suitable for learning long-distance text information, and then improve to get the long and short term memory neural network(LSTM). Due to the complex structure, the computational parameters and more and more computationally intensive of LSTM neural network, there comes the GRU model, which is simpler than LSTM and has fewer parameters. In order to further reduce the model training time, improve the accuracy and reduce the loss rate, researchers proposed the BiGRU-Attention model, which can reduce the computational effort of the model and fully extract the feature information of the text context compared to a single hybrid model of LSTM or GRU neural network [4]. However, in effectiveness of text classification and information extraction, the complexity of the original sentence makes the model not as effective as it could be. If dividing the original sentence into several word vectors and then merging them into sentence vectors, the classification and extraction effect will be greatly improved. Therefore, this paper incorporates the RoBERTa [12] model and designs and applies a RoBERTa-BiGRU-Attention fusion model.



Figure 1: RoBERTa-BiGRU-Attention fusion model

3.1.2.Text classification based on Dual Contrastive Learning

Because the deep learning model network is deep and needs a large amount of data, and the data set used in this paper is limited, it may be difficult to achieve the best results; this paper introduced dual contrastive learning. Dual Contrastive Learning is a new learning framework. In unsupervised learning tasks, contrastive learning has been proved effective in characterizing downstream tasks and achieving good results [5]. The contrastive learning approach can also be applied to supervised learning, but the supervised contrastive learning approach lacks principled application and reduces representation validity compared to traditional supervised representation learning, which requires developing another classification algorithm to solve the classification task.



Figure 2: A text classification framework using Dual Contrastive Learning

Using Roberta model as encoder f, obtaining each token feature of the sequence, splicing the labeled text with the input text with [SEP] and fusing the original position vectors, text vectors, and word vectors in the model. where $e_{relevant}$ and $e_{irrelevant}$ are classifier representations and e_{CLS} is the feature representation. After DuaCL training, the positive samples keep approaching while the negative samples keep moving away.

3.2. Tourism Product Heat Ranking

Since the travel guide text is unstructured text, it is necessary to extract the valid entities from the travel guide text. Then do sentiment analysis on the sentence in the travel guide where the entity is located and evaluating and ranking the heat of tourism products each year based on the analysis results.

3.2.1.BERT-BiLSTM-CRF named entity identification

This paper takes a deep learning based approach, The BERT-BiLSTM-CRF model is an end-to-end deep learning model developed based on the BiLSTM-CRF model without manual feature induction, which can fulfill the current needs of Chinese address parsing and address element annotation tasks [6]. This model from the bottom up consists of an encoder, a BiLSTM neural network layer, and a conditional random field (CRF) layer: The encoder is a character-level Chinese BERT-based model, which maps the input Chinese address characters into a low-dimensional dense real number space, and mines the potential semantics embedded in each type of address element in the Chinese address; The BiLSTM neural network layer takes the character vector transformed from the encoder as input and captures the forward (left-to-right) and backward (right-to-left) bi-directional features of the Chinese address sequence; The conditional random field layer takes the bi-directional features extracted from the upstream BiLSTM as input, and combines the Bioes labeling paradigm to generate the labels corresponding to each character in the address, so as to further parse the Chinese address into various address elements according to the labels.

using the adversarial training approach [7], as shown in Figure 3. During the training process, first BERT will generate initial vectors from the input text, and then add some perturbations on it to generate adversarial samples as variants of the original samples, which are easily misleading to the model. The initial vectors and the adversarial samples will be fed together into BiLSTM for training, during which the neural network will learn more robust parameters to resist the adversarial sample attack.



Figure 3: Using Adversarial Training in BERT-BiLSTM-CRF Models

3.2.2.A multidimensional heat evaluation model based on the improved Wilson interval method

When processing the evaluation data of the sample, the traditional heat analysis algorithm based on user voting has obvious shortcomings: Delicious algorithm simply ranks by the number of users' comments per unit of time, ignoring comment emotion; Reddit sorting algorithm simply takes the absolute value of the difference between positive and negative reviews as the depth of affirmation, regardless of the positive rating; The traditional Wilson interval sorting algorithm works well in solving small samples, but lacks the consideration of the problem that product heat decays as time goes on. For this reason, this paper proposed an improved time factor-incorporating algorithm that using the lower bound of the confidence interval to replace the favorable rating by introducing and improving the Wilson confidence interval estimation.

The Wilson score interval correction formula proposed by Wilson [8]:

$$\frac{p + \frac{1}{2n} z_{1-\alpha/2}^2 \pm z_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n} + \frac{z_{1-\alpha/2}^2}{4n^2}}}{1 + \frac{1}{n} z_{1-\alpha/2}^2}$$
(1)

In the formula, p denotes the proportion of the sample rated as good; n denotes the number of samples; $z_{1-\frac{\alpha}{2}}^2$ denotes the statistic corresponding to a certain confidence level and is a constant, for example, the statistical value of z is 1.96 at 95% confidence level. Then calculates the heat score based on the lower bound of formula (1).

When n is large enough, formula (2) tends to \hat{p} . Since the score calculated by formula (2) is a number between (0, 1), the ranking can be based on the lower value of this confidence interval; the higher the value, the higher the ranking. Also considering the user's browsing, commenting and time factors of the information, defining the calculation formula for the product heat analysis algorithm based on user comments as:

$$R = \frac{\log_{10} W_{views} + W_{score}}{(W_{age} + 1)^G}$$
(2)

3.3.Local Tourism Graph Construction

3.3.1.Association rule mining based on improved Apriori algorithm

Obtain the entities set in each travel guide by named entity identification technology, there are redundant identical items between sets and it is difficult to find the association between point sets. The Apriori algorithm can find the association items to get the relationships between tourism entities. The improved Apriori algorithm [9] only needs to traverse the database once to obtain the association rule results between frequent item sets. The main steps to improve the Apriori algorithm are as follows:





Step 1: Delete irrelevant transaction records.

Let the total number of transaction items be m and the traversal database be D. When D_x (x=1, 2, ..., m). count=1, delete D_x , the number of deleted transaction items is counted as 1, and so on after the traversal loop to get the new database D'. Let the set of interest be B. If D'_x (x=1, 2, ..., n), B $\notin D'_x$, then delete D'_x and traverse the loop to get the new data set D".

Step 2: Mine the frequent item sets.

Counts each transaction item to obtain the candidate 1-item set, where the items greater than or equal to min_sup will form the frequent item set L_1 . Self-connect the generated frequent item set L_1 to generate the candidate 2-item set, and perform the set intersection operation to obtain the transaction TID set, where the items greater than or equal to min_sup will form the frequent item set L_2 . Compute the modulus $|L_k|$ of L_k and end the operation when $|L_k| \leq k$ to obtain the frequent item set L. Otherwise repeat step B.

Step 3: Mine association rules.

Calculate the degree of support and confidence, analyze the association relationship between variables, summarize certain regularity between variables and generate association rules, the process is shown in Figure 6.

3.3.2.Implicit relationship discovery based on GNNLP model

Since the improved Apriori algorithm can only identify frequent item sets from known relations and

mine known associated edges but cannot predict unknown missing edges, the constructed graphs are not complete with node relations when constructing the knowledge graph. For this reason, this paper proposes the GNNLP model. After generating the knowledge graph, adopt neural network function to nonlinearly fit the nodes in the graph, and fulfill the aggregation and update of the node information in the graph by GNN-related algorithm to convert the Maoming tourism knowledge graph into a GNN graph with neural network.

The aggregation operation collects information at the neighbors of each node by means of an aggregation function, set the aggregation function aggregate(x), where x denotes aggregating the information from all neighboring nodes of the target node [10].

$$a_v^{(k)} = \text{aggregate} \ {}^{(k)}(\{h_u^{(k-1)}: u \in N(v)\})$$

$$(3)$$

 $a_v^{(k)}$ denotes the kth aggregation result of a node, N(v) denotes the neighbor nodes of node v, and $h_u^{(k-1)}$ denotes the k-1th state of the aggregation of neighbor node u. Different functions are suitable for different graph structures.



Figure 5: GNNLP model process

Update Process, perform a specific operation between the result after information aggregation and the central node as the initial state of the node in the next layer (i.e., update the hidden state of the node). Set the update function combine(y), where y denotes a specific operation between the result of the previous step of aggregation and the target node.

$$a_v^{(k)} = \text{aggregate } {}^{(k)} \left(\left\{ h_u^{(k-1)} : u \in N(v) \right\} \right)$$

$$\tag{4}$$

 $h_v^{(k)}$ denotes the kth update result of node v, and $h_v^{(k-1)}$ denotes the k-1th state of node v. Once repeat the above operation, the number of layers of the neural network adds 1. Keep aggregating and updating until the number of updates reaches 1. Then dividing the nodes in the GNN graph into subgraphs according to the number of paths and distances between different node pairs; then calculate the path similarity and node similarity between different node pairs respectively, fuse and process the two to obtain the final link similarity between node pairs; finally, ranking according to the final link similarity, then perform graph neural network link prediction to fulfill the discovery of implicit relationships between nodes, the GNNLP model process is shown in Figure 7.

Figure 6 is a local tourism graph constructed by visualization techniques based on the mining results of the improved Apriori algorithm, on this basis, using the GNNLP model constructed in this paper to discover the implicit relationships between nodes, the result is obtained as shown in Figure 7, where the blue bolded edges represent the newly discovered relationships between nodes after passing the GNNLP model.

4. Experiment and Analysis

In order to verify the rationality of the model constructed in this paper, the following validation experiments are designed.

4.1.Text classification results and analysis

On the basis of the introduction and analysis above, this paper divides the training set and test set in the ratio of 4:1, trains and tests the commonly used text classification models and the RoBERTa-BiGRU-Attention fusion model and RoBERTa-DualCL model used in this paper. The effects of each model are shown in Table 3:

Comparison and evaluation of text classification results of typical models				
	Model	Loss	Acc(%)	
	RNN	0.2937	91.632	
	LSTM	0.2455	91.894	
	Text CNN	0.2234	91.793	
	Text RNN	0.2256	91.833	
	BERT	0.1940	92.534	
	Roberta	0.1672	92.976	
	Roberta-BiGRU	0.1340	93.112	
	Roberta-BiGRU-MA	0.0913	93.572	
	Roberta-DuaCL	0.00186	96.900	

Table 3

From the above results, the Roberta-DuaCL model with dual contrastive learning has the highest accuracy, classifying 1312 correctly in 1354 test sets, achieving a correct rate of 96.90%. Using the Dual Contrastive Learning framework for data enhancement achieves better results on small samples, so the model can be used to classify texts.

Using the Roberta-DuaCL model to classify tourism texts, the results showed 4315 texts in the tourism-related category and 1971 texts in the tourism-unrelated category.

4.2.Named entity identification results and analysis

There is no standard for entities in the tourism field, and most of the existing naming identification tasks in the tourism field are only for attraction identification and cannot meet the needs of this topic. This paper carefully analyzes the travel guide data and defines 6 entities in the tourism field following the principle of each entity type can completely cover the entities in the tourism field and has no intersection according to the task requirements: SCENIC, HOTEL, DIET, ENTERTAINMENT, CULTURE and VILLAGE. The model obtained after training and optimizing with the constructed named entity identification dataset in the tourism field works well for entity recognition, extracting totally 2246 entities from the travel guide.

Example of named entity identification results

Travel Guide ID		Entity Identification Results	Entity Type	Publish Time
1267	1252	Opencast Mine Good Lake Ecopark	SCENIC	2021-04-0818:33
70	1009	One Piece Eggplant	DIET	2019-02-1821:12
1799	1066	Mixing Powder	DIET	2019-08-2621:28

2396	1190	Rubber tube	ENTERTAINMENT	2020-08-02 11:16
3113	1239	Dragon Head Mountain	SCENIC	2021-02-07 15:04
1883	1099	White cut chicken	DIET	2019-09-30 21:21
2147	1161	Fantasy Crystal Church	SCENIC	2020-04-24 15:57
2516	1197	Hot spring area	ENTERTAINMENT	2020-08-21 22:30

Rank the products by their heats, the ranking results of the top-ranked tourism products are shown in Figure 7:

Table 5

Top-ranked tourism products in heat

Pr	oduct	Product		Heat of	
ID		Types	Name of Product	Product	Year
0	ID5	DIET	Youzhipin Pastry	1	2018
1	ID35	DIET	Hello Fried Chicken (Fangxing)	1	2020
2	ID62	DIET	Wheat Crust Pastry (Development Zone)	1	2021
3	ID361	SCENIC	Romantic Coast	1	2019
4	ID24	DIET	CAKE Love in the Black Forest	0.990349	2019
5	ID2	DIET	Qing Xiang Bakery(Che Tian Street)	0.965898	2019
6	ID22	DIET	Fruits of the Degree (Weimin Store)	0.962907	2019
7	ID23	DIET	Get Together Time (Guangming Store)	0.925987	2019
8	ID2	DIET	Qing Xiang Bakery(Che Tian Street)	0.92179	2018

4.3. Results and analysis of relation extraction

Some association relationships mined by the improved Apriori algorithm are shown in Table 6 below:

Table 6

association relationships between some of the tourism products

Product 1	Product 2	Relevance	Association Type
Shuidong mustard	Dredging powder	0.80	DIET——DIET
Chicken heart yellow skin	jackfruit	1.00	DIET——DIET
Street steak	Wyndham hotel	0.67	DIET——HOTEL
Eat your mouth out	Yushuigu hot spring hotel	0.50	DIET——HOTEL
Cockscomb stone	Fangji island	0.67	SCENIC——SCENIC
Great horn bay	Ten-Mile silver beach	0.75	SCENIC——SCENIC

Based on the strong association rules mined by the improved Apriori algorithm, 11 implied highlevel association concepts were predicted by the GNNLP model in 2018 and 2019. For example, Fangji island and Seaview Bay Hotel are linked through the upper concept Fangji island tourist area, Hailing island and dredging powder are linked through the upper concept hailing island Ten-Mile silver beach scenic spot 7 implied high-level association concepts were predicted in 2020 and 2021, romantic coast and lobster are linked through the upper concept Wyndham Hotel, Opencast Mine Good Lake Ecopark and Shijue temple are linked through the upper concept Opencast Mine This leads to the implied high-level concept, for example, the connection between Fangji island and Seaview Bay Hotel through Fangji island can lead to the inference that tourists tend to stay in sea view hotels by the coast when visiting Fangji island, which conclusion can promote the development of the surrounding hotels and B&Bs.



Figure 7: Enhanced and completed tourism graph after link prediction

5. Concluding remarks

This paper uses natural language processing data mining methods to analyze the development of surrounding travel of the city during the COVID-19 epidemic by building a local tourism graph; based on 2 core technologies: Dual Contrastive Learning text classification and graph neural network, solves 4 problems of WeChat public article classification, surrounding travel tourism product heat analysis, local tourism graph construction and analysis and change analysis of tourism product demand before and after the epidemic; based on traditional models, improved designs RoBERTa-BiGRU-Attention fusion model, Dual Contrastive Learning, BERT-BiLSTM-CRF named entity identification technique, improved Apriori algorithm, GNNLP model and other models and methods; demonstrates the rationality and efficiency of the improved model through comparative tests; essentially overcomes the shortcomings and loopholes of the traditional model and achieves a satisfactory result.

The results show that the method adopted in this paper and the improved model algorithm both achieve good results. Firstly, they solve the problem of data decentralization and fragmentation, improving the accuracy of text classification; secondly, they extract the relevant tourism elements from the text clearly and accurately, enhance the comprehensiveness and accuracy of heat analysis; finally, they fulfill the deep-level mining of the implied high-level concepts and the weak relationships obtained from prediction can enhance and complete the original graph, construct a knowledge graph with reference significance to the development of local travel in the circumstance of the epidemic.

6. References

- [1] Zhang Ju, Feng Ao, Zhang Xuelei et al. A Sentiment Analysis Method for Travel Text Fused with Text-Rank [J]. Computer Science and Applications, 2022, 12
- [2] Cui Liping, Gulila Adonbek, Wang Zhiyue. Named entity identification in tourism field based on directed graph model[J]. Computer Engineering, 2022, 48(2)
- [3] Zhang, Nuo. Research on Knowledge Graph Construction Method for Shanxi Tourism [D]. Shanxi University.
- [4] Cai Wenxing, Li Xingdong. Sentiment analysis of scenic spot reviews based on BERT model[J]. Journal of Guizhou University (Natural Science Edition), 2021,38(2):57-60. DOI:10.15958/j.cnki.gdxbzrb.2021.02.11.
- [5] Niu T, Xiong C, Socher R. Deleter: Leveraging BERT to Perform Unsupervised Successive Text Compression[J]. 2019.
- [6] Zhao P, Sun LY, Wan Y et al. Named entity identification of Chinese attractions based on BERT+BiLSTM+CRF[J]. Computer System Applications, 2020, 29(6):6.
- [7] Cao Liujuan, Kuang Huafeng, Liu Hong et al. Geometric constrained adversarial training with twolabel supervision[J]. Journal of Software,2022,33(4):1218-1230. DOI:10.13328/j.cnki.jos.006477.
- [8] Xu Linlong, Fu Jiansheng, Jiang Chunheng et al. A ranking algorithm of product favorability based on Wilson interval[J]. Computer Technology and Development,2015(5):168-171. DOI:10.3969/j.issn.1673-629X.2015.05.040.
- [9] Liu Wenya, Xu Yongneng. Subway fault association rule mining based on improved Apriori algorithm[J]. Journal of Arms and Equipment Engineering,2021,42(12):210-215. DOI:10.11809/bqzbgcxb2021.12.033.
- [10] Wu, Guodong. Research on personalized item recommendation based on deep learning [D]. Shanghai: Donghua University,2020.
- [11] Li Jiahui. Research on multi-domain text classification methods based on RoBERTa and cyclic convolutional multi-task learning [D]. Harbin Institute of Technology.