ORIGINAL ARTICLE

Sentiment analysis algorithm using contrastive learning and adversarial training for POI recommendation

Shaowei Huang[1] · Xiangping Wu[1,2] · Xiangyang Wu[3] · Ke Wang[1,2]

Abstract

Finding a suitable POI based on the user's needs and intentions is a complex decision-making process. Obtaining valuable information from the vast amount of social media data and using it for travel recommendations is a challenging issue. Traditional POI recommendation algorithms do not fully take into account the true feelings of customers about tourist attractions implied in social media data because they usually require a large amount of tagged travel commentary data. This study presents an aspect-based sentiment analysis model and POI recommendation method to accurately capture sentiment information contained in social media data with a small amount of tagged data. The pre-training model BERT is used to obtain the embedded representation of words that fuse the semantic information of the text. Using contrastive learning, point clusters belonging to the same class in the embedded space of words are pulled together, and sample clusters from different classes are separated. The potential relationship between comment ratings and their impact on user perception is analyzed, and the best performance formula for the loss function is determined. The test accuracy and F1-Score of the model in the experiment improved by 13.03% and 12.23%, respectively, compared to the BERT base model. POI recommendation validation is performed using a variety of recommendation algorithms. The experimental results show that the addition of aspect-based sentiment attributes can effectively improve the accuracy of recommendations.

Keywords Sentiment analysis · Contrastive learning · POI recommendation · BERT

✉ Xiangping Wu
xiangpingwu@cjlu.edu.cn

Shaowei Huang
1259721632@qq.com

Xiangyang Wu
wuxy@hdu.edu.cn

Ke Wang
wangke@cjlu.edu.cn

1 College of Information Engineering, China Jiliang University, Xiasha, Hangzhou 310026, Zhejiang, China

2 Key Laboratory of Electromagnetic Wave Information Technology and Metrology of Zhejiang Province, China Jiliang University, Xiasha, Hangzhou 310026, Zhejiang, China

3 School of Computer Science and Technology, Hangzhou Dianzi University, Xiasha, Hangzhou 310026, Zhejiang, China

1 Introduction

With the rapid development of mobile internet and explosive growth of social media users, social media data is rapidly increasing in scale, category, diversity, and so on. A large amount of data is generated every day (Toivonen et al. 2019). Relevant scholars have opened up new research fields for information processing problems of massive social media data (Dogan and Birant 2021). For instance, when planning their holidays, travelers tend to consult experiences and opinions of other tourists on various network platforms (Shafqat and Byun 2019). The vast amount of public data that can be used to conduct user surveys and tap potential customers is thus used as part of market research by the travel industry. It is impractical to manually process a large volume of user-generated data. Some data mining tools and algorithms have been proposed to help users extract relevant information from large data. Obtaining comment texts from the internet and building a recommendation framework can prove to be beneficial for tourism professionals and users. The

tourism recommendation system based on sentiment analysis (Fayyaz et al. 2020) is a very new research topic, and its great potential in the tourism industry has attracted researchers. However, user reviews may mention different categories, such as POI locations, features, and services, which can make it difficult to extract useful information from the data. Additionally, there are many factors that contribute to the complexity of the study, such as differences among users and how reviewers write their comments. For the same travel POI, different customers may provide feedback in completely different ways, describing what they are interested in and making decisions based on their preferences.

To address the above-mentioned challenges in the field of tourism recommendation, this paper proposes an aspect-level sentiment analysis based on the Supervised Contrastive Learning (SCL) model of Bidirectional Encoder Representation from Transformers (BERT) (Devlin et al. 2018). This model classifies sentiment attributes and attributes of the comments and analyzes the potential relationship between comment ratings and comment text in the context of online travel text, and their impact on user perception. The extracted sentiment information at the potential level can be effectively utilized in the travel recommendation system to improve the accuracy of the recommendation.

This paper presents an aspect-based sentiment analysis model for tourism recommendations. The model delves deeply into aspect-based sentiment attributes in POI comment texts, calculates the POI sentiment feature matrix and the user sentiment feature matrix, and generates travel POI recommendations based on tourists' historical comments and visits. Compared with the baseline method, the incorporation of aspect-based sentiment attributes utilizes both the textual appearance and semantic features in the comments, and considers both the tourists' preferences and the POI characteristics to enhance the accuracy of the recommendations.

1. The BERT-CL aspect absed sentiment analysis model is proposed. The BERT-based word embedding representation innovatively uses contrast learning to aggregate clusters of similar points in the word embedding space, while separating clusters of samples from different classes. It also uses adversarial training to generate positive and negative samples for contrast learning, which improves the performance under small sample conditions.
2. The potential relationship between comment ratings and text in text data and the impact on user perception are analyzed. The optimal loss function is determined.
3. Sufficient experiments are conducted to demonstrate the effectiveness of the proposed model.

2 Related works

Sentiment analysis is a technique used in natural language processing, computational linguistics and text mining to determine the sentiment implied in a text. In recent years, researchers have proposed a variety of sentiment analysis models for human-computer interaction (Xu et al. 2023; Wang et al. 2022; Tsai et al. 2022), information retrieval, and multimodal signals (Cambria et al. 2017) in different domains, such as tourism, finance, shopping, etc. Hu et al. (2022) demonstrates the feasibility and rationality of the joint modelling of emotion and emotion from a psychological point of view. The authors propose a unified framework for multimodal knowledge sharing that introduces cross-modal comparative learning and combines auditory and visual modal representations with multilevel textual features. Zhao et al. (2021) proposed a BERT-based sentiment analysis and key entity detection method for financial analysis, which is used to analyze online financial texts. Using the pre-training model, sentiment analysis is performed first, then key entity detection is performed as a sentence matching or machine reading task, and integrated learning is used to improve the performance of the model. Bhat et al. (2020) attempted to analyze the emotions of various texts extracted from different social media sites in response to the COVID-19 outbreak on social media. Manguri's model (Manguri et al. 2020) sentiment analysis of recent tweets about the global COVID-19 outbreak found on the social media site Twitter. In addition, Ruz et al. (2020) performs sentiment analysis on Twitter data during critical events using a Bayesian network classifier.

Due to the recent surge in the number of online communities, the analysis of online comments, mainly expressed by consumers, and comments on specific products, has attracted the attention of many researchers. Many recent models have been used by researchers to analyze these comments (Gu and Zhang 2022). Some examples of such studies are as follows: Amplayo and Song (2017) summarized the adaptive aspect-based sentiment analysis of several short online reviews collected from rotten tomatoes (movie reviews), Amazon (shopping product reviews), etc. Abdi et al. (2018) introduces an adaptive aspect-based sentiment analysis that summarizes several brief online comments. The proposed model utilizes an online movie dataset. The development of machine learning and deep learning has greatly promoted the progress of effective analysis. And proposed a sentiment analysis system based on machine learning to analyze travelers' comments on Egyptian hotels. This study presents a traveler comment sentiment classification algorithm, analyses travelers' comments on Egyptian hotels, and classifies each mood according to hotel characteristics. Emotion models use

three classification techniques: Support Vector Machine, Naive Bayesian, and Decision Tree. The Naive Bayesian classifier provides the highest accuracy of 85%, accuracy of 0.61, and recall rate of 0.67. Alsubari et al. (2022) proposed a system for collecting comments from the Web and created a structured overview of the comments so that they could be easily accessed. For each sample, the classifier produces positive or negative polarity values. Hybrid systems utilize a combination of statistical classifiers and polarity classifications from information extraction systems. The pre-training language model (Devlin et al. 2018) represented by BERT has been widely used to fine-tune downstream NLP tasks. The cross-entropy loss (Rumelhart et al. 1986) is used to fine-tune the pre-training model, which calculates the KL divergence between the label's One-Hot vector and the model output prediction, and then predicts using a linear classifier. However, there are also shortcomings in this standard process. Liu et al. (2016) pointed out that loss of cross-entropy may result in poor generalization performance (Cao et al. 2019), and may lack robustness to noise labels (Zhang and Sabuncu 2018). In addition, there is some instability in the cross-entropy loss function during fine-tuning BERT under small sample conditions (Dodge et al. 2020) By adding directly to the pre-training model, it may face the problem of over-fitting, especially when the training data is limited (Zhang et al. 2020).

In contrast to regular recommendation scenarios, the preferences of users in the travel recommendation field often show distinct characteristics in terms of time and geographical location. Online travel planning is a tedious task and each region has a large number of points of interest. This often raises questions. For example, advertisements usually do not reveal their own shortcomings in order to attract users. As a result, feedback from visited users is more trustworthy, and many researchers are trying to design systems that address this task, so that users in need can more easily access useful information from comments. Many technicians have built various comment-based recommendation systems using different technologies, such as sentiment analysis of consumers' online comments on online websites. Chen et al. (2021) In order to achieve personalised, interactive and traffic-aware travel planning, a dynamic point of interest network model is built based on a combination of location-based social network (LBSN) and taxi GPS digital footprint. To improve the representation of user and POI features, Zhang et al. (2022) proposed an information dissemination mechanism to model high-order connections of different POI-related relationships. Kim et al. (2021), uses two types of information derived from check-in history as input for factorising: a transition pattern between two POIs and the number of visits to POIs.

Combining the research status of sentiment analysis and travel recommendation, the existing methods of sentiment analysis-based recommendation generally use statistical-based text analysis methods, such as topic modeling and TF-IDF. They take the content of user comments as an important reference for the recommendation, which shows greater improvement than the method of recommendation relying solely on access records. However, traditional text analysis methods cannot effectively mine contextual information, semantic information, and domain knowledge implied in the text. They also pay more attention to the impact of time and space factors on users, while ignoring the differences between individual users (Fig. 1).

3 Method

In the field of tourism recommendation, the aspect-based sentiment analysis model shown in Fig. 2 is utilized to extract potential user features and POI (point of interest) features from tourism review data. The primary issue with the sentiment analysis model is the insufficient availability of accurately labeled data sets. To address this, a tourist attraction comment data set containing tags has been prepared. Since some comments may express diverse sentiments, the sentiment attributes are categorized into five levels. Positive and negative sentiments are each divided into two levels, respectively, to indicate the degree of positivity or negativity.

Word vector coding uses a pre-trained BERT model. First, regular expressions are used to denoise the comment data. For Chinese comment data, it is necessary to remove punctuation, emoticons and word segmentation. Only the characters with semantic information in the sentence are left. For the maximum coding length of the BERT word vector, we have counted the sentence length distribution in the travel comments and blog data, as shown in Fig. 1, and selected max lenth as 128. It is fed into the model to generate word vectors, and uses a transformer and self-attention mechanisms
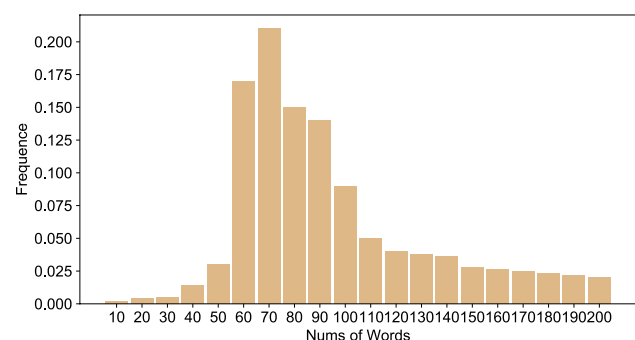

Fig. 1 A plot of token frequency versus the number of tokens for a review

Fig. 2 POI recommendation method based on aspect-based sentiment analysis
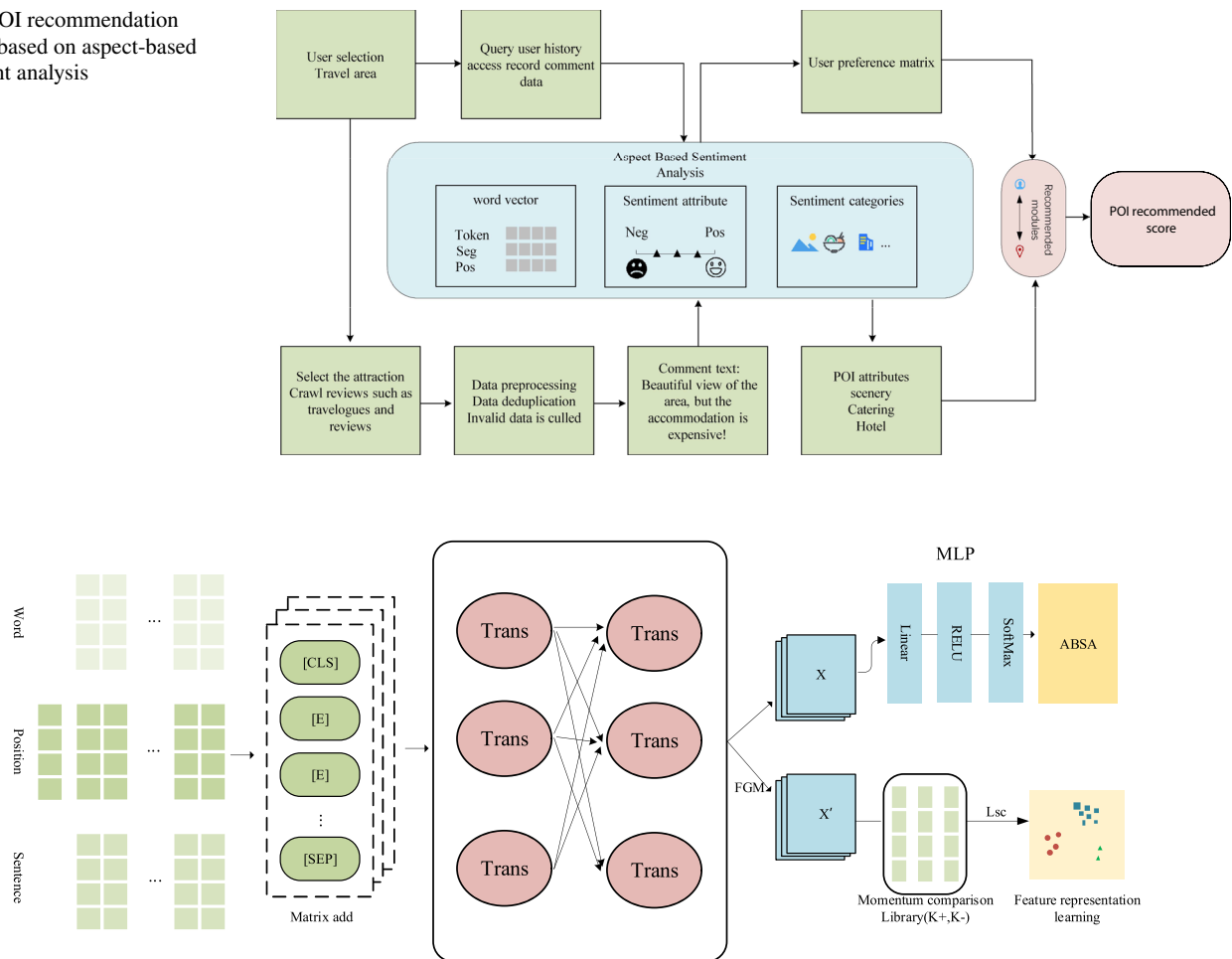

Fig. 3 BERT-CL aspect-based sentiment analysis model framework

to mine the potential connections between words. The word embedding process is shown in Fig. 3. The input e_i to the transformer is formed by adding three vectors representing different sentence features. The first vector is the [cls] flag, which can be used for downstream classification tasks. And the [sep] flag at the end of the sentence is used as a separator between different sentences.The word embedding vector is the static encoding of words. The position vector encodes the position of the participle in the sentence. in the sentence. The maximum number of words per sentence is 126, after leaving room for two fixed characters. Sentences longer than this are truncated to ensure consistency in the dimensions of the word vector.

3.1 BERT-CL model

A neural network model is proposed to perform aspect-based sentiment prediction. The model is based on the pre-trained language model BERT and named BERT-CL. The model includes a coding module, a supervised contrastive

learning module for optimizing coding features, and a training module for expanding the number of positive and negative samples.

The structure of the BERT-CL model combining BERT and contrast learning is shown in Fig. 3. The model uses adversarial training to generate contrast learning samples. It consists of an encoding module, a supervised comparison learning module and an adversarial training module, respectively. First, the BERT language model is used to obtain a vector of representative words containing semantic information. Then, the FGM module is used to generate adversarial samples. Then, the supervised contrast learning module is used to extract the similarities between the same categories and the differences between different categories. Finally, the sentiment attributes of the samples are output.

3.2 BERT encoder

Word vector coding uses a pre-trained BERT model. The comment data is input into the model to generate word

vectors, and the transformer combined with the self-attention mechanism is used to mine the potential connections between words. The word embedding process is shown in Fig. 4. The input of the Transformer is formed by adding three vectors representing different sentence features. The first vector is the [cls] flag that can be used for downstream classification tasks, and the [sep] flag at the end of the sentence is used as the separator of different sentences; Word embedding vector is the static encoding of words; Position vector PE records the position of word segmentation in a sentence, which is different from embedding words in English. In Chinese, each character is embedded as a word.

The input representational word vectors are passed through the multilayer attention model of BERT, and the output matrix of the last layer is denoted as $T_r = T_1, T_2..., T_r$. The T_r matrix row dimension is the same as the BERT input matrix. Each row vector represents the

unambiguous depth vector of the subword, which is used as input for the downstream task.

3.3 Contrastive learning for feature extraction

For natural language processing tasks, supervised contrastive learning methods for supervised learning tasks such as text classification (Gunel et al. 2020). Gunel et al. (2020). In order to obtain the word vector representation with stronger generalisation performance in sentiment analysis tasks, the model is based on the training strategy of contrastive learning. on the training strategy of contrastive learning. As shown in Fig. 5, the goal of contrastive learning is to construct a compact representation cluster within the same class, while maintaining the distance between clusters of different classes. A supervised contrastive learning framework based on category is introduced using the label information from the downstream task dataset to
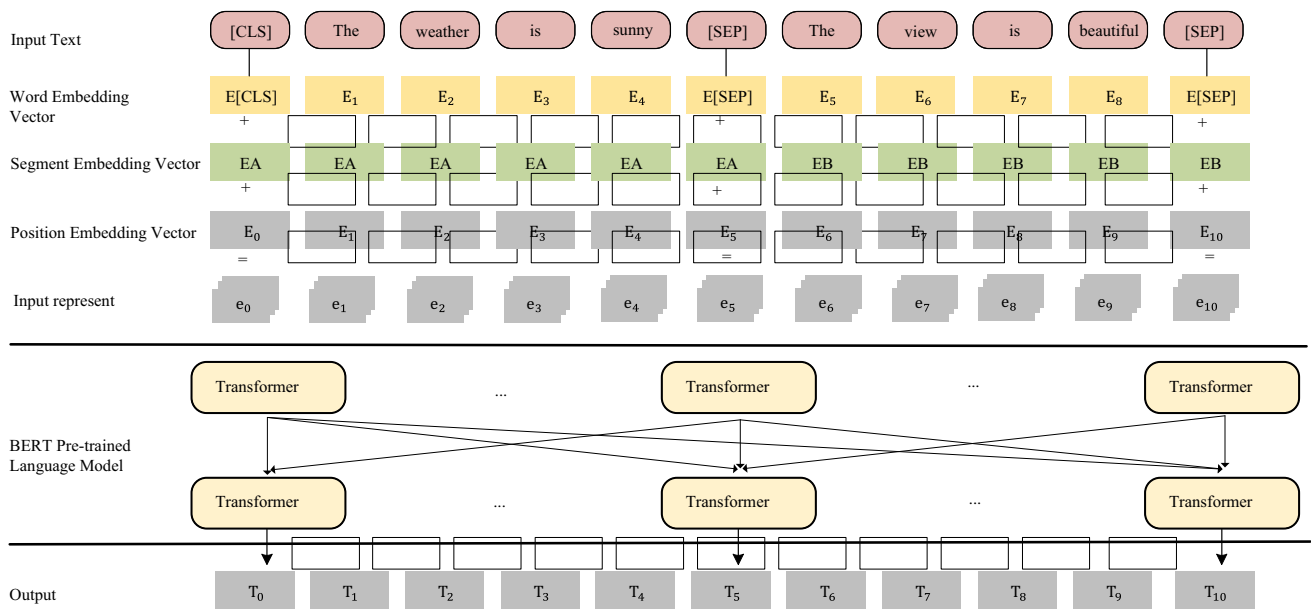

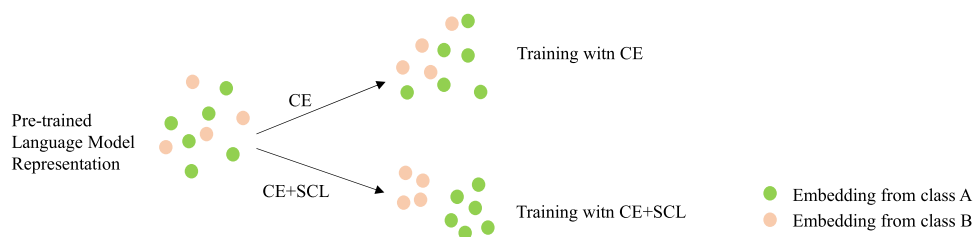**Fig. 4** BERT model word embedding process

Fig. 5 Contrastive learning aims to push examples of the same class closer and examples of different classes farther away. Take the binary classification task as an example, where category A is a negative review and category B is a positive movie review. This loss usually applies to any multi-class classification setup

cluster the representations. Unlike the traditional contrastive learning framework, we do not use the finite expansion-based method to construct positive samples, but use the affirmative words of classes extracted from the same class of a given example to construct affirmative words. Compared to the expanded affirmative words, this kind of affirmative words are relatively more abundant and useful, and their semantics can also be diversified.

Contrastive learning (Chen et al. 2020) is a training strategy based on the similarity of samples in different categories. It has been widely used in a variety of deep learning models (Yin et al. 2017). Comparison loss mainly includes noise comparison estimation loss (Mnih and Kavukcuoglu 2013) and N-Pair loss (Sohn 2016). The contrast learning framework is widely used in self-supervised tasks (Chen et al. 2020). In supervised scenarios, contrast loss can also be used to modify the loss function. The supervised contrastive learning loss is added to the normal training as an additional task, and the parameter updating process of the model is still based on the linear classifier. The purpose of comparison loss is to capture the similarity between samples of the same category and compare them with samples of different categories. For effective analysis problems with C categories, the data set will contain N training samples, $\{xi, yi\}, i \in (1, N)$. Y_i is a dataset with the same label Y_i total number of samples; The build loss function is:

$$L = (1 - \lambda)L_{\text{CE}} + \lambda L_C \tag{1}$$

Contrastive learning trains a representation (q) with positive samples (k_+) and negative samples (k_-). The similarity of samples is determined by dot product, and the contrastive loss with multiple positive and negative samples is expressed as:

$$L_C = -\log \frac{\sqrt{\frac{q \cdot k_+}{\tau}}}{\sum_{k_i \in \{k_-, k_+\}} \sqrt{\frac{q \cdot k_+}{\tau}}} \tag{2}$$

Where τ is the temperature super parameter, $\{k_-, k_+\}$ is the sum of more than one positive sample and negative sample, a total of n samples. This form of the loss function is closely related to the widely used information noise comparison estimation (Oord et al. 2018). It is widely used in self-supervised contrastive learning tasks (Sohn 2016).

As shown in the formula (2), multiple positive and negative samples are used in calculating the contrastive loss. Therefore, the dynamic contrastive framework is used to update the positive and negative samples to better represent the classification.

The disadvantage of contrastive loss is that a large number of negative samples are required. Chen et al. (2020)

introduced a dynamic contrastive method, which combines a set of data samples, and the encoders of these samples are updated during the dynamic process. The end-to-end update process uses samples from the current small batch, so it is limited by the memory size of the GPU. The repository method stores representations that cannot be updated in real time. The dynamic contrast method makes it possible to train multiple positive and negative supervised contrastive losses in large-scale pre-training language model tuning.

In the process of contrastive learning training, combining negative samples can be used to improve the sampling effect of continuous high-dimensional space under coding representation. Therefore, the dynamic comparison framework (MOCO) uses a queue-based update strategy to handle a large number of negative samples. In the dynamic comparison framework, there are two independent encoders: query encoder and key value encoder. The query encoder is updated by using the gradient descent of the query sample. The optimization of the key encoder is solved by using a dynamic process from querying the parameters of the encoder, as follows:

$$\theta_k \leftarrow m\theta_k + (1 - m)\theta_q \tag{3}$$

Here θ_q and θ_k are encoders that update only θ_q during gradient updates by back propagation.

The negative sample representation vector is first pushed into the recursive queue, and after the key encoder is updated based on the dynamic process of the query encoder, the samples at the end of the queue are updated only by encoding with the key encoder. Through the dynamic updating process, the contrastive learning process can handle a large number of positive and negative samples, because it does not need to calculate the gradients of all positive and negative samples.

Different from the traditional dynamic contrast learning frameworks used in the field of computer vision, there are many ways to construct large numbers of positive and negative samples in the field of NLP. Push all these samples into the training queue and construct positive and negative samples based on the label of the query samples.

3.4 Adversarial training training for improving small sample classification performance

Aiming at the problems of difficult data annotation and limited manual annotation accuracy in the field of text classification, the method of confrontation training is used in the training process to improve the robustness and generalization ability of the model under small sample training. Countermeasure training is a kind of countermeasure defense. In the training, countermeasure samples constructed according

to certain rules are added to the original data set to enhance the robustness of the model to countermeasure samples, so as to improve the prediction accuracy of the model. Confrontation training is defined as:

$$\min_{\theta} E_{x,y} \sim D\left[\max_{\Delta x \in \omega} L(x + \Delta x, y; \theta)\right] \tag{4}$$

$$\Delta x = \epsilon \nabla x L(x, y; \theta) \tag{5}$$

Where D represents training set, x represents input, y represents label, θ Is the model parameter, $L(x + \Delta x, y; \theta)$ Is the loss of a single sample, Δx is the adversarial disturbance and Ω is the disturbance space.

The process of confrontation training:

1. Adding a position embedding vector to each tag to specify its position in the sequence; The word embedding, position embedding and sentence embedding vectors are combined into a vector of $3 \cdot Maxlen \cdot 768$.
2. According to the size of the batch size in the training process, the combined vector obtained in N steps is input at a time, and the whole input sequence passes through the transformer module, as shown in Fig. 2.
3. Add adversarial disturbance to the word vector of each sample Δx, it is expressed as:

$$\Delta x = \min_{\theta} E_{(x,y) \sim D[\max_{\Delta x \in \Omega} L(x+\Delta x, y; \theta)]} \tag{6}$$

Where d represents the whole training set, X is the combination vector obtained in step 3, and Y corresponds to the label of each sentence in X, including aspect label and sentiment label, θ Represents model parameters, $l(x, y; \theta)$ is the loss function of a single sample, Δx is the adversarial disturbance and Ω is the disturbance space. Δx satisfies the following constraints to make its salient characteristics almost the same as the initial sample:

$$\|\Delta x\| \leq \epsilon \tag{7}$$

take Δx for Standardization:

$$\Delta x = \epsilon \frac{\nabla_x L(x, y; \theta)}{\|\nabla_x L(x, y; \theta)\|} \tag{8}$$

After construct the counter sample $x + \Delta x$ for each sample, use $(x + \Delta x, y; \theta)$ minimize the loss update parameter as a training data pair θ; And as a sample of the same category of contrastive learning, it was added to the training. The robustness and generalization ability of the model under small sample training are improved by using confrontation training method for the input vectors.

Experiments on aspect-based sentiment analysis tasks and robustness tests show that by combining contrastive learning and confrontation training with the traditional fine-tuning process, the overall performance of the model can be effectively improved under the condition of small samples, and the robustness to confrontation attacks can be improved.

The deep neural network based on BERT-CL and the fast gradient method (FGM) training method improve the classification accuracy of the model in small samples. FGM, as a method of defending against adversary attacks, adds the generated adversary samples to the training set for data enhancement, so that the model can learn the adversary samples while training. Taking the word vector as the input of the aspect-based sentiment analysis model, the sentiment data of the sample is shown in Table 1, which means that the score range of the sentiment attributes of the sample is 1–5. The larger the score, the more positive the sentiment attributes of the sample are, and vice versa. Thus, the evaluation vector r of each comment is obtained $S = \{r_1, r_2, r_i, ..., r_n\}$, where $0 \leq R_i \leq 9.0 \leq s_i \leq 4$. Respectively represent the sentiment attribute value of each aspect of the comment.

4 Experiment and result analysis

4.1 Data

Our data includes travel-related text data from Ctrip and Qunar.com. Each data contains several fields with different meanings. These include basic information of POIs provided by social media and user comment text. The dataset contains 26,820 instance samples. 2140 of them are labelled samples. Each sample contains the following fields:

The purpose of preprocessing is to convert some disordered and noisy data into data suitable for computing. Massive datasets pose a huge challenge to data cleaning because manual editing datasets is impractical and inefficient. Therefore, data processing is inseparable from the support of automation software; There are many kinds of data processing software, including various programming languages, application software packages of various data processing methods, etc. The python language is used for data processing. The collected semi-structured raw data containing repeated information and noise information is transformed into the

Table 1 Travel blog text data description

The field name	Illustrate
User number	ID of each comment user
Release date	Date the comment was posted
Name of POI	Relevant attractions mentioned in the comments
Comments text	User generated comment text data
Pype	Self driving/Tour group
Position	Location of tourist attractions
Number of likes	Number of likes for this comment
Score	Overall evaluation score given by the user

Table 2 Text data sentiment categories

Cate	Nums
Natural scene	322
Cultural scenery	275
Environmental health	240
Service	217
Experience	354
Restaurant	228
Hotel	185
Shopping	363
Entertainment	237
Popularity	215
Degree of recommendation	238

Table 3 Text dataset affects the distribution of attributes

Sentiment	Nums
0	71
1	165
2	1230
3	510
4	160

correct form required for subsequent analysis steps through data cleaning, extraction, transformation and fusion.

In the process of data tagging, the open data set on the internet and the self-built data set in the tourism field are used to tag each data with sentiment attributes. The tags are divided into five levels: very dissatisfied, dissatisfied, intermediate attitude, satisfied, and very satisfied. As described by Wang et al. (2020), the advantage of five-level tags over the two-level sentiment attributes of 'positive' and 'negative' is that they can more accurately mine users' sentiments. Inspired by the document (Ray et al. 2021), we divide the attribute categories of POI into 11 aspect-based categories such as 'cultural scenery' and 'environmental sanitation' to describe the characteristics of POI. Labels including sentiment aspects include scenery, catering, hotel, shopping, entertainment, and the proportion of each aspect in the total sample, as shown in Table 2.

After categorising and sentiment analyzing the comment data, calculate the aspect level sentiment score matrix for each user or POI:

$$S_a = \sum_{i \in (0,N)} k_i lgN \tag{9}$$

S_a in Formula A indicates the sentiment score to be calculated for each goal, k_i denotes the i_{th} comment contained by the object, N denotes the number of comments on a topic in a user's or POI comment, and the higher the model's

sentiment score for one to five levels, the more positive it is (Table 3).

4.2 Hyperparameter setting and evaluation

The performance of the model is evaluated using the following metrics, where accuracy, recall, and F1-Score are used to evaluate the performance of the model sentiment analysis, and NDCG and RMSE are used to evaluate the performance of the recommendation algorithm.

Accuracy:

$$Accuracy(f;D) = \frac{1}{n} \sum_{i=1}^{n} (f(x_i) = label_i) \tag{10}$$

Recall:

$$Recall = \triangleq \frac{TP}{TP + FN} \tag{11}$$

F1 score: In order to be able to refer to both precision and recall, the index F1 score is introduced, taking into account precision and recall:

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precesion + Recall} \tag{12}$$

Recommended tasks typically use NDCG to evaluate sorted results, which are formulated as:

$$NDCG_k = \frac{DCG_k}{IDCG_k} \tag{13}$$

where:

$$DCG_k = \sum_{i=1}^{k} \frac{rel_i}{\log_2(i + 1)} \tag{14}$$

and:

$$IDCG_k = \sum_{i=1}^{|REL|} \frac{2(rel_i) - 1}{\log_2(i + 1)} \tag{15}$$

RMSE (Root-Mean-Square Error) can be used to assess the deviation between predicted and true values in the recommendation system expressed as:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{k} (Y_i - f(x_i))^2} \tag{16}$$

4.3 Sentiment analysis

In order to extract more valid features, the fine-tuning process differs from that exemplified by previous text

classification tasks. Instead of using the [CLS] vector of the BERT model for sentiment analysis, the model uses a pooling of each word vector contained in the sentence for feature extraction. The model normalizes the feature representation (Yin and Shang 2022), because the normalization method has been widely used in the contrastive learning method, and has been proved to be useful by the experimental results.

In order to use the label information more directly, the model adds the contrastive loss and the original cross entropy loss in the fine-tuning process. The BERT model uses a fixed-length sequence. BERT pre-training model version used is Chinese-BERT-wwm (Devlin et al. 2018). And sets the maximum length of comment statements to 128, and the batch size is set to 32, Adam optimizer is used. The development environment is TensorFlow 1.4. The program is run in the Linux operating system and trained with GTX1080ti. The data set was divided into 7:2:1 for training, verification, and testing.

The process of obtaining aspect-based sentiment evaluation of POI is completed by two stages: model training and deployment. In the training stage of the model, the marked data is input into the BERT-CL model for training. In the deployment phase of the model, the tourism comment data is input into the model to obtain the aspect-based sentiment attributes of each sentence in each comment.

Compared with the original BERT model, the addition of comparative learning and antagonistic training methods plays a key role in improving the accuracy of the model. Table 4 shows the performance improvement of BERT-CL over the baseline model in the tourism datasets described in Sect. 4.1. BERT combines several traditional neural network structures such as RNN, CNN, and LSTM as baseline models for comparison. Compared with the feature extraction method of the traditional model, contrastive learning has achieved significant accuracy improvement in the fine-tuning stage of the BERT model. At the same time, we further

Table 4 Precesion and F1 scores of different models in the tourism review data set

Model	Precision (%)	Recall (%)	F1 scores
BERT	64.35	65.13	64.26
BERT-RNN	67.24	67.19	67.36
BERT-CNN	73.79	73.33	73.24
BERT-BiLSTM	70.59	70.47	70.28
BERT-CL(no AT)	75.94	75.68	75.84
BERT-CL(with AT)	77.38	76.94	76.03

explored the improvement of adversarial training on model performance. Table 4 shows the change in the accuracy of the classification model before and after adding the adversarial training method. It can be observed that the accuracy of the model is further improved after the addition of the adversarial training method.

The validity of the model is verified by using tourism data sets and multiple public data sets, and the construction standards of tourism data sets and the sources of public data sets are explained.

The results of the sentiment analysis method of BERT-CL are compared with some of advance methods of sentiment analysis (Yin et al. 2017), such as recurrent neural network (RNN), convolution neural network(CNN) and bidirectional LSTM (BiLSTM). At the same time, ablation experiments are set to verify the effectiveness of introducing contrastive learning and confrontation training to improve the performance of the model. Tables 4 and 5 list the precisio of RNN, CNN and BiLSTM, respectively. It can be seen from Table 5 that the BERT-CL model produces better results for all classes than advance model. Figure 6 shows the comparison of the BERT-CL model with the baseline model and the BERT-CNN model. In the training process, the convergence speed of BERT-CL is faster, and the final loss is less than other
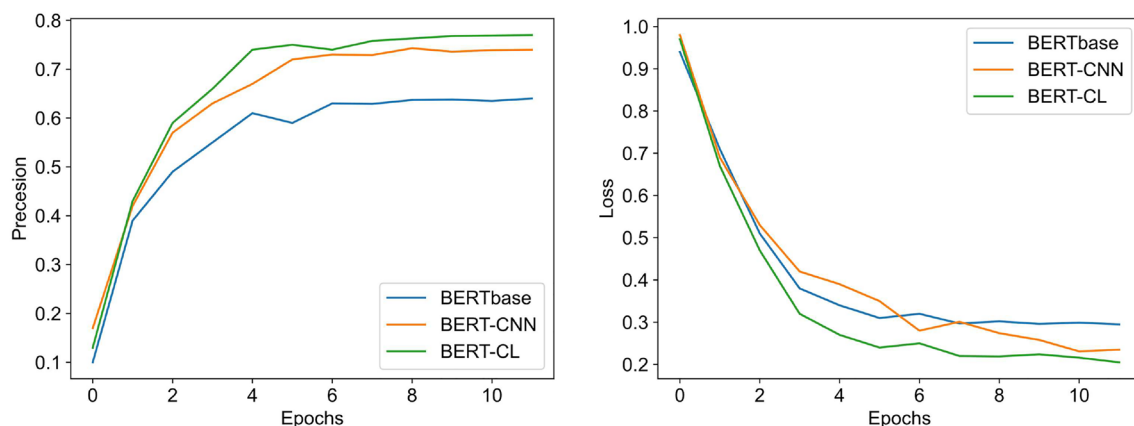

Fig. 6 Comparison of training process with baseline model and BERT-CNN model

Table 5 Precesion of different models in public dataset

Model	SST-2	IMDB
BERT	92.68	64.35
BERT-RNN	92.43	67.24
BERT-CNN	92.72	73.79
BERT-BiLSTM	92.26	70.59
BERT-CL(no AT)	92.70	75.94
BERT-CL(with AT)	92.74	76.81

Table 7 Few-shot learning test result on SST-2(precision%)

Model	5%	3%	1%	0.5%
BERT	74.25	72.47	69.52	64.26
BERT-RNN	77.56	76.72	74.73	73.28
BERT-CNN	83.67	80.72	78.83	75.83
BERT-BiLSTM	80.64	78.67	75.89	72.22
BERT-CL(no AT)	85.57	84.25	82.76	80.78
BERT-CL(with AT)	87.63	87.72	87.61	85.72

models. The test results of each training Epoch also show that the model has better fitting ability in the tourism data set.

There is a sample imbalance problem in the sentiment analysis dataset in the tourism field. The sample size of different sentiment attributes varies widely. As shown in Table 3, the number of samples with a neutral affective attribute is significantly more than the positive and negative samples. To explore the performance of the model prediction of sentiment properties under the condition of few-shot and sample imbalance. We list the model's precesio, recall rate, and F1 score for samples of different affective attributes in Table 6. We can observe that the model's very negative and very positive predictions for a small sample size are slightly lower than those of the neutral sample. It shows that BERT-CL has strong robustness.

In Table 7, we report the few-shot learning results in SST-2 data set, in which only training samples with a minimum training set size of 0.5% are used, respectively. We observed that in all data sets, our BERT-CL significantly improved the performance compared with baseline, and compared with BERT, BERT-CNN increased by 11.46 and

9.91 percentage points, respectively, indicating that the emotion analysis model we proposed is effective on different data sets. And with the gradual reduction of training samples, he decline of accuracy of BERT-CL model is much smaller than that of the comparison model.

In order to further verify the performance of the model in predicting noisy data, the reverse translation model commonly used in NLP is used to add noise to the test data. This paper conducts experiments on tourism data and SST-2 data sets. Table 8 records the ACC of different models under different degrees of noise interference. The experimental setup followed the training method described in Sect. 4.2. T is a parameter used to adjust the amount of noise in the back translation model. The larger the value of T, the more noise is added to the data. We observed that in all noise data, when the noise level is low ($T = 0.3$), all models, including the original BERT model, remain unchanged. This shows that BERT has strong robustness in semantic understanding. When the noise level gradually increases, BERT-CL improves significantly compared with other models, especially in travel and IMDB data sets.

In Fig. 7, the TSNE diagram of embedded learning representation on SST-2 test set was shown when training with 1% training samples. Compare the BERT baseline model with BERT-CL. It is clear to see that adding contrastive learning causes the same-labelled samples to be more compactly clustered. The embedding distribution learned by BERT is close to a random distribution. Figure 8 shows the effect of the contrast learning temperature hyperparameter on the model, where the temperature determines the intensity of separation of different classes in contrast learning. The effect of τ on accuracy is tested in different data sets.
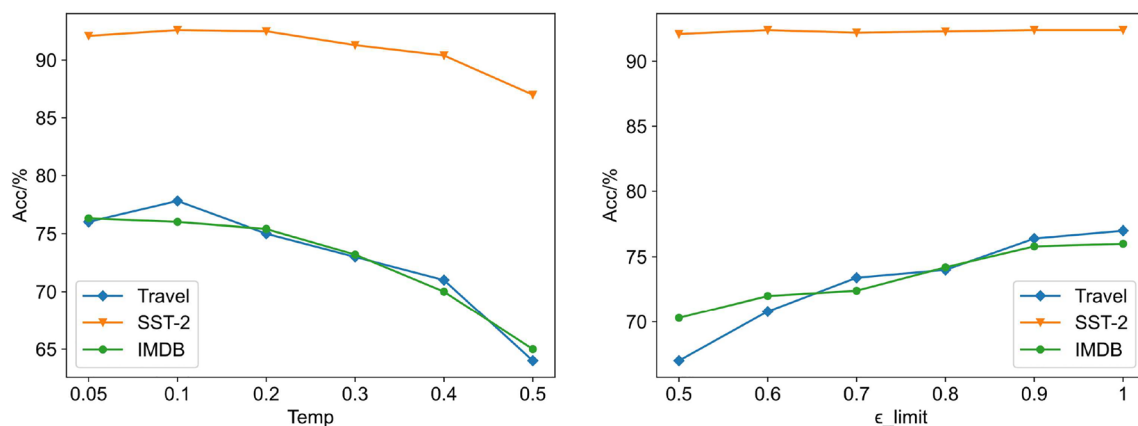
4.4 POI recommendations

The architecture of the tourism recommendation system is shown in Fig. 2. Firstly, the unstructured data is crawled and preprocessed. And the structured data set is extracted from the online travel review website. Then category sentiment features are extracted from the datasets. Classification of sentiment attributes and sentiment aspects of comments using the BERT-CL model.

Table 6 Precesion of different models in travel dataset

Model	Sentiment attribute	Precision (%)	Recall (%)	F1 score
BERT-RNN	0	65.94	64.46	64.27
	1	67.13	66.38	66.67
	2	67.45	67.16	66.52
	3	67.21	66.62	66.27
	4	66.97	66.82	65.23
BERT-CNN	0	70.47	69.65	69.35
	1	73.24	73.38	72.72
	2	74.17	73.83	74.02
	3	73.93	74.18	73.73
	4	71.34	70.83	70.82
BERT-CL	0	75.28	74.83	74.37
	1	76.43	75.34	75.42
	2	77.63	77.03	76.27
	3	77.18	76.31	76.95
	4	75.27	74.73	74.82

Table 8 Results on three datasets for robustness across noisy augmented training sets(precision%)

Data	Model	Original	$T = 0.3$	$T = 0.5$	$T = 0.7$	$T = 0.9$
Travel	BERT	63.45	62.72	61.02	61.56	60.72
	BERT-CL	75.94	75.72	74.83	74.72	72.87
	BERT-CL(AT)	77.38	76.62	76.78	75.83	73.62
SST-2	BERT	92.68	92.76	91.72	92.42	91.82
	BERT-CL	92.70	92.26	92.72	92.83	92.43
	BERT-CL(AT)	92.74	92.65	92.87	91.98	91.35
IMDB	BERT	64.35	62.72	62.76	60.72	57.82
	BERT-CL	75.94	72.83	70.17	68.32	64.32
	BERT-CL(AT)	76.81	75.62	72.27	71.73	69.65
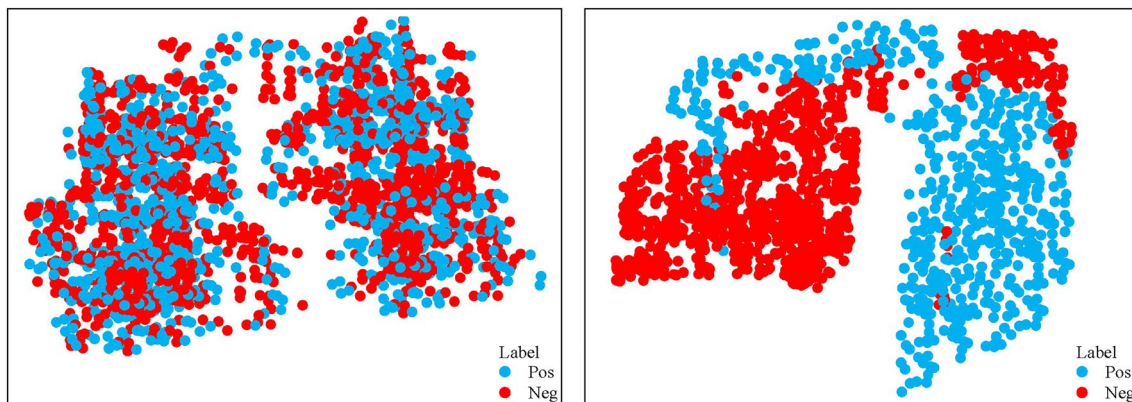
Fig. 7 In a few-shot training setup fine-tuned using 1% SST-2 samples during the training phase, the learned tSNE diagram of the CLS embedded SST-2 test set compares the BERT baseline model to our proposed BERT-CL sentiment analysis model (right). Blue: positive example; Red: Negative example (Color figure online)

Fig. 8 Performance (Acc) comparisons on different numbers of learn temperature hyperparameters and adversity limits on different datasets

Multiple recommendation algorithm models are used for experimental verification. Multiple recommendation algorithm models are used for experimental validation. Fine-grained sentiment attributes information of users' check-in information and comments are used for POI recommendations. To verify the effectiveness of fine-grained sentiment attribute information on recommendation

Table 9 Using different recommendation algorithms, the addition of aspect-based sentiment scores effectively improves the recommendation success rate

Recommend algorithm	Precision (%)	RMSE	Precision (%)*	RMSE*
LLORMA	13.45	0.87	14.17	0.81
CF-NADE	15.52	0.84	15.96	0.79
GC-MC	16.92	0.83	18.14	0.75

*Means that the BERT-CL sentiment score feature is used

accuracy improvement. The models used are as follows. The model used is as follows:

1. LLORMA (Lee et al. 2016) is a matrix decomposition model using local low-dimensional submatrix decomposition.
2. CF-NADE (Zheng et al. 2016) replaced the characteristics of RBM with neural autoregressive distribution estimator (nade) for rating reconstruction.
3. GC-MC (Berg et al. 2017) is a graph-based AE framework, which applies GNN to binary interaction graph for rating link reconstruction.

Qunar.com's travel blog records which POIs have been visited by tourists. the recommendation algorithm validation experiments use the data of tourists' visit history from Go.com to evaluate the accuracy of model recommendations. please note that we desensitize all the data to protect the implicit of users. In the experiment, the data are divided into training, validation and test data according to 7:2:1. For users with access records, their historical posting data are read and their comment texts are analyzed using a fine-grained sentiment model to obtain the user preference value S_{ua}. In the experiments, the three recommendation algorithms use the same dataset, with the difference that the three algorithms use the sentiment scores generated by the BERT base model and the more accurate sentiment scores generated by BERT-CL, respectively, with the model The parameter settings are the same as in Sect. 4.2.

Table 9 shows that using different recommendation algorithms, the addition of fine-grained sentiment attributes of BERT-CL effectively improves the recommendation success rate compared to the sentiment scores generated by the BERT base model, where the GC-MC algorithm has the best performance improvement when combined with the recommendations generated by BERT-CL, and the addition of fine-grained sentiment scores improves the accuracy of the model by 1.23%, and indicates that the deviation value of the model-generated Meanwhile, LLORMA and CF-NADE also have different degrees of improvement.

Table 9 shows that using different recommendation algorithms, the addition of aspect sentiment attributes can effectively improve the recommendation success rate. Table 9 shows that the GC-MC algorithm has the best performance improvement in the recommendation generated by combining BERT-CL. The addition of BERT-CL structure improves the accuracy of the model by 1.23%. Moreover, the RMSE representing the deviation between the recommended and real data generated by the model decreased by 0.05. At the same time, LLORMA and CF-NADE also have different degrees of improvement.

In summary, the BERT pre-training model combined with contrast learning effectively improves the accuracy of the recommendation system, and the input vector of BERT is mapped from the task corpus in a large corpus collection instead of being learned from the task corpus. This allows the vectors trained by the model to have a higher degree of generalization.

5 Conclusion

Travel recommendation systems can help users make the best travel decisions by providing travel suggestions personalized to their preferences. Compared with existing recommendation methods based on social media sentiment analysis, the new approach applies an improved BERT deep learning network. Firstly, reviews are classified into different aspects of category. Secondly, the reviews are sentiment analyzed. Finally, appropriate travel POIs are recommended for the users based on their input preference. The experiments use the travel POI reviews dataset collected by Qunar.com. For the sentiment analysis task, an accuracy rate higher than other advanced models is achieved. Adding fine-grained sentiment attributes as additional attribute values to the recommendation system effectively improves the accuracy of recommendations. Bert pre-trained model also has shortcomings. It mainly focuses on the signals within the original language, and the overall semantics of the acquired semantic knowledge units may be biased. This shortcoming is more obvious when training Chinese corpus. The BERT pre-training model does not use the masking mechanism when fine-tuning downstream tasks. This results in data bias in the training phase and the model usage phase. This will have some impact on the actual prediction results. In future research, the rapid iterative update of social media data is addressed. Studying the change of sentiment attributes of comments over time may help to improve the accuracy of recommendations. And consider using the optimization model of BERT and new pre-training models such as XLNET to analyze more multi-sentiment tendency text. The richer sentiment tendency is obtained.

Author contribution SH and XW wrote the main manuscript text, XW and KW prepared the experiment for Sect. 5. All authors reviewed the manuscript.

Declarations

Conflict of interest The authors declare no competing interests.

References

Abdi A, Shamsuddin SM, Hasan S, Piran J (2018) Machine learning-based multi-documents sentiment-oriented summarization using linguistic treatment. Expert Syst Appl 109:66–85

Alsubari SN, Deshmukh SN, Alqarni AA, Alsharif N, Aldhyani TH, Alsaade FW, Khalaf OI (2022) Data analytics for the identification of fake reviews using supervised learning. Comput Mater Continua 70(2):3189–3204

Amplayo RK, Song M (2017) An adaptable fine-grained sentiment analysis for summarization of multiple short online reviews. Data Knowl Eng 110:54–67

Berg RVD, Kipf TN, Welling M (2017) Graph convolutional matrix completion. arXiv preprint arXiv:1706.02263

Bhat M, Qadri M, Noor-ul-Asrar Beg MK, Ahanger N, Agarwal B (2020) Sentiment analysis of social media response on the Covid19 outbreak. Brain, Behavior, and Immunity 87:136

Cambria E, Das D, Bandyopadhyay S, Feraco A (2017) Affective computing and sentiment analysis. Pract Guide Sentim Anal 1–10

Cao K, Wei C, Gaidon A, Arechiga N, Ma T (2019) Learning imbalanced datasets with label-distribution-aware margin loss. Adv Neural Inf Process. Syst 32

Chen T, Kornblith S, Norouzi M, Hinton G (2020) A simple framework for contrastive learning of visual representations. In: International conference on machine learning pp 1597–1607

Chen C, Zhang D, Wang Y, Huang H (2021) TripPlanner: personalized trip planning leveraging heterogeneous trajectory data. In: Enabling Smart Urban Services with GPS Trajectory Data. Springer, Singapore, pp 225–247

Devlin J, Chang MW, Lee K, Toutanova K (2018) BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805

Dodge J, Ilharco G, Schwartz R, Farhadi A, Hajishirzi H, Smith N (2020) Fine-tuning pretrained language models: Weight initializations, data orders, and early stopping. arXiv preprint arXiv:2002.06305

Dogan A, Birant D (2021) Machine learning and data mining in manufacturing. Expert Syst Appl 166:114060

Fayyaz Z, Ebrahimian M, Nawara D, Ibrahim A, Kashef R (2020) Recommendation systems: algorithms, challenges, metrics, and business opportunities. Appl Sci 10(21):7748

Gunel B, Du J, Conneau A, Stoyanov V (2020) Supervised contrastive learning for pre-trained language model fine-tuning. arXiv preprint arXiv:2011.01403

Gu P, Zhang Z (2022) Dual-attention based joint aspect sentiment classification model. In: Web Engineering: 22nd International Conference, ICWE (2022) Bari, Italy, July 5–8, 2022, Proceedings. Springer International Publishing, Cham, pp 252–267

Hu G, Lin TE, Zhao Y, Lu G, Wu Y, Li Y (2022) UniMSE: towards unified multimodal sentiment analysis and emotion recognition. arXiv preprint arXiv:2211.11256

Kim JS, Kim JW, Chung YD (2021) Successive point-of-interest recommendation with local differential privacy. IEEE Access 9:66371–66386

Lee J, Kim S, Lebanon G, Singer Y, Bengio S (2016) LLORMA: Local low-rank matrix approximation

Liu W, Wen Y, Yu Z, Yang M (2016) Large-margin softmax loss for convolutional neural networks. arXiv preprint arXiv:1612.02295

Manguri KH, Ramadhan RN, Amin PRM (2020) Twitter sentiment analysis on worldwide COVID-19 outbreaks. Kurd J Appl Res 54-65

Mnih A, Kavukcuoglu K (2013) Learning word embeddings efficiently with noise-contrastive estimation. Adv Neural Inf Process Syst 26

Oord AVD, Li Y, Vinyals O (2018) Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748

Ray B, Garain A, Sarkar R (2021) An ensemble-based hotel recommender system using sentiment analysis and aspect categorization of hotel reviews. Appl Soft Comput 98:106935

Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. Nature 323(6088):533–536

Ruz GA, Henríquez PA, Mascareño A (2020) Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers. Future Gener Comput Syst 106:92–104

Shafqat W, Byun YC (2019) A recommendation mechanism for under-emphasized tourist spots using topic modeling and sentiment analysis. Sustainability 12(1):320

Sohn K (2016) Improved deep metric learning with multi-class n-pair loss objective. Adv Neural Inf Process Syst 29

Toivonen T, Heikinheimo V, Fink C, Hausmann A, Hiippala T, Järv O, Di Minin E (2019) Social media data for conservation science: a methodological overview. Biol Conserv 233:298–315

Tsai YH, Chang CM, Chen KH, Hwang SY (2022) An integration of TextGCN and autoencoder into aspect-based sentiment analysis. In: Big Data Analytics and Knowledge Discovery: 24th International Conference, DaWaK (2022) Vienna, Austria, August 22–24, 2022, Proceedings. Springer International Publishing, Cham, pp 3–16

Wang Z, Ho SB, Cambria E (2020) Multi-level fine-scaled sentiment sensing with ambivalence handling. Int J Uncertain Fuzziness Knowl Based Syst 28(04):683–697

Wang T, Sun B, Tong Y (2022). Auto-ABSA: automatic detection of aspects in aspect-based sentiment analysis. arXiv preprint arXiv:2202.00484

Xu L, Pang X, Wu J, Cai M, Peng J (2023) Learn from structural scope: improving aspect-level sentiment analysis with hybrid graph convolutional networks. Neurocomputing 518:373–383

Yin W, Kann K, Yu M, Schütze H (2017) Comparative study of CNN and RNN for natural language processing. arXiv preprint arXiv:1702.01923

Yin W, Kann K, Yu M, Schütze H (2017) Comparative study of CNN and RNN for natural language processing. arXiv preprint arXiv:1702.01923

Yin W, Shang L (2022) Efficient Nearest Neighbor Emotion Classification with BERT-whitening. In: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing pp 4738-4745

Zhang S, Bai Z, Li P, Chang Y (2022) Multi-graph convolutional network for fine-grained and personalized POI recommendation. Electronics 11(18):2966

Zhang Z, Sabuncu M (2018) Generalized cross entropy loss for training deep neural networks with noisy labels. Adv Neural Inf Process Syst 31

Zhang T, Wu F, Katiyar A, Weinberger KQ, Artzi Y (2020) Revisiting few-sample BERT fine-tuning. arXiv preprint arXiv:2006.05987

Zhao L, Li L, Zheng X, Zhang J (2021) A BERT based sentiment analysis and key entity detection approach for online financial texts. In 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD) (pp. 1233-1238). IEEE

Zheng Y, Tang B, Ding W, Zhou H (2016) A neural autoregressive approach to collaborative filtering. In: International Conference on Machine Learning pp 764–773