ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

Natural Language Processing in tourism Review sentiment Analysis Tourist satisfaction Prediction Model Construction

Lu Tang[1] *and Dandan Sun*[2,*]

[1] College of Humanities and Tourism,Wuhu Vocational Technical University, Wuhu 241006, Anhui, China

[2] Department of Business Administration, Shandong Institute of Commerce and Technology, Jinan 250103, Jinan, China

SUMMARY: *Tourism review texts reflect tourists' feelings on scenic spots, hotels, transportation, catering and services, and also provide a calculation basis for satisfaction prediction. This paper integrates vocabulary normalization, contextual semantic coding, polarity recognition, aspect extraction and rating regression to construct a Natural Language Processing (NLP) based travel review sentiment analysis and satisfaction prediction network. This study selected 48,620 English travel reviews to complete text cleaning, label mapping and context embedding. The model uses BERT to extract semantic representation, and uses BiLSTM to depict the before and after semantics. Combined with attention weighting, multi-dimensional semantic fusion and consistency calibration, the model outputs three types of emotion labels and five levels of satisfaction scores. The experiments were completed in Python 3.10 environment, and the training, validation and test sets were divided into 7:1:2. The results show that the accuracy rate of the model is 94.18%, the precision rate is 93.47%, the recall rate is 92.86%, the F1-score is 93.15%, the AUC is 0.962, and the MAE is 0.218, which is better than the baseline models such as TF-IDF-SVM, CNN, BiLSTM and BERT. It can provide data support for tourism review understanding and tourist experience analysis.*

KEYWORDS: *Natural language processing; Travel review; Sentiment analysis; Satisfaction prediction model*

1 Introduction

The online travel platform has saved a large number of tourist reviews, and the experience of scenic spot visit, hotel stay, transportation connection, catering consumption and after-sales service is recorded as ratings and emotional statements. Such texts are large in quantity and semantic dispersion, and it is difficult to form a stable judgment by manual reading. Natural language processing can transform unstructured reviews into computable semantic vectors, so that sentiment polarity and satisfaction score enter the same prediction link. Puh and BagićBabac studied the application of machine learning in the prediction of sentiment and rating of tourist reviews, and showed that there is a modelable relationship between review semantics and rating results [1]. Ameur et al. reviewed sentiment analysis methods of hotel reviews and pointed out that preprocessing and feature expression would affect recognition results [2]. Ali et al. proposed LDA topic sentiment analysis method, which combined comment topic and sentiment tendency [3].

*fixxluckyy@163.com
https://doi.org/10.65102/is2026820
Published: 30 April, 2026

Travel reviews do not only contain simple judgments of "satisfied" or "dissatisfied". Tourists often describe environment, price, queue and service in the same comment, and there are overlapping and inversion of semantic polarity. Nawawi et al. adopted aspect-level sentiment analysis and zero-shot learning to explore tourist experience in online reviews, so that service evaluation could be reduced to dimensions [4]. Zakarija et al. constructed the Dubrovnik tourist experience sentiment model based on the Tripadvisor platform, indicating that regional context can affect sentiment classification results [5]. Hanić et al. compared the performance of various machine learning models in sentiment analysis and rating prediction of restaurant reviews, and proved that the model structure would change the fitting accuracy of sentiment and rating [6]. These studies provide a basis for tourism review calculation, but a tighter mapping mechanism for context dependence, implicit attitude and satisfaction rating still needs to be established.

This paper constructs an NLP deep network model around tourism review sentiment analysis and tourist satisfaction prediction. The model takes online travel reviews as input, and after text cleaning, word segmentation, stop word filtering, word form normalization and abnormal symbol processing, the review sequence suitable for neural network training is generated. Charfaoui and Mussard proposed a machine learning sentiment analysis path for tourism insights, proving that computational models are able to extract interpretable information from touriser-generated content [7]. George and Ramos studied the content sentiment analysis generated by tourists in Kangyang tourism destinations, and showed that the review text could reflect the service perception and spatial feeling [8]. This paper introduces a pre-trained language model to extract semantic features, combines BiLSTM to depict the semantic dependence before and after, and then uses the attention layer to calculate the weights of sentiment words, aspect words and degree words, so that the satisfaction prediction is associated with the interpretable semantic dimension.

This paper designs a sentiment analysis and satisfaction prediction model for tourism reviews. The data set contains 48620 English tourism reviews, covering four types of scenes: scenic spots, hotels, transportation and comprehensive services. The review text is labeled as positive, neutral and negative sentiment, and is mapped to five-level satisfaction labels according to the score. Diaz-Pacheco et al. proposed a deep learning method for tourism text topic discovery and sentiment analysis, showing that deep networks are suitable for complex tourism semantic recognition [9]. Collini et al. studied data-driven scenic spot evaluation and tourist visit prediction, and integrated reputation assessment and visit prediction into the same computational framework [10]. In this paper, the sentiment classification task and satisfaction regression task are jointly modeled, and a shared semantic encoding layer is used to reduce feature fragmentation. The cross-entropy loss and mean absolute error loss are used to jointly constrain the model training.

Aiming at the task of tourism review sentiment analysis and tourist satisfaction prediction, this paper constructs a deep semantic modeling framework based on natural language processing, which integrates review cleaning, semantic encoding, sentiment polarity recognition and satisfaction prediction into the same calculation process. Instead of separating sentiment classification and rating prediction, the framework uses a shared semantic representation layer to establish the association between review text, sentiment orientation and satisfaction level, so that the model can simultaneously capture tourists' expression differences on services, environments, prices, transportation and facilities. This paper further introduces the attention weight mechanism to weight the semantic segments that affect the satisfaction judgment, so that the prediction results have a clearer source of features. This model provides a reusable calculation method for tourism platform review understanding, quantitative analysis of tourist experience and intelligent processing of service evaluation.

2 Literature Review

Focusing on tourism review sentiment analysis and tourist satisfaction prediction, existing research has been carried out from the directions of semantic mining, deep classification, rating fitting and user-generated content modeling. Eshkevari et al. constructed an end-to-end ranking system based on customer reviews, which combined semantic mining with multi-criteria decision making method to calculate object ranking relationship from text reviews. This study proved that review semantics could be transformed into rankable computational features, but its focus was more on ranking result output. The linkage between emotional polarity and satisfaction level is rarely expressed [11]. Ganji et al. proposed a deep learning sentiment classification method for hotel recommendation system, and introduced data balance processing to improve the influence of class sample distribution difference on model training. This method is suitable for processing accommodation evaluation texts with uneven proportion of positive and negative reviews, but the expression of fine-grained semantic dimensions such as service, environment and price in reviews is still limited [12]. Roy studied the relationship between travelers' online reviews and hotel performance, combined the accommodation theory with sentiment analysis, and showed that the review text could reflect the hotel business performance and tourists' perception differences, which provided a scene basis for satisfaction prediction. However, the calculation link was still dominated by emotional interpretation, and a unified neural network structure for rating prediction was lacking [13].

Hidayati proposed an aspect-level hotel review sentiment analysis method that combines latent Dirichlet allocation (LDA) with machine learning algorithm. The service dimension in reviews is extracted through topic discovery, and then sentiment judgment is made, so that a clear correspondence between text topic and sentiment tendency is established [14]. Marutho and Rustad further proposed a combination of sentence embedding Transformer, Bayesian search clustering and sparse attention mechanism to improve aspect-level sentiment analysis, which can enhance the context retention ability of semantic representation and make the model's weight allocation to comment segments more stable [15]. Jeong and Lee used ChatGPT to conduct an aspect level review analysis of hotel service errors, and used large language models for service error identification and text classification, showing the applicability of generative language models in review understanding. However, the model output still needs to be constrained by controllable feature weights and prediction targets [16].

Malik and Bilal systematically sorted out the research path of natural language processing analysis of online customer reviews, and proposed a classification framework from data collection, text cleaning, feature representation, emotion recognition to evaluation application, which provided a complete technical reference for tourism review processing [17]. Wankhade et al. reviewed the methods, applications and challenges of sentiment analysis, covering dictionary methods, machine learning, deep learning and hybrid models, indicating that sentiment recognition has shifted from shallow word frequency statistics to semantic context modeling [18]. Mustak et al. used machine learning to develop customer insights from user-generated content and demonstrated that review data can be transformed into computational insights for behavior understanding and service judgment [19]. Hartmann et al. studied the accuracy and application boundaries of sentiment analysis, and emphasized the differences in the results of different models in context transfer, domain vocabulary and real applications, suggesting that tourism review satisfaction prediction needs to balance classification accuracy, semantic interpretability and rating mapping stability [20].

In order to further sort out the technical emphasis of existing research in text semantic modeling, sentiment classification, aspect recognition and satisfaction prediction, Table 1 compares the relevant literature from four aspects of calculation method, data object, technical contribution and application boundary.

Table 1: Comparison of models related to sentiment analysis and satisfaction prediction for travel reviews

Reference	Computational Method	Data Object	Technical Contribution	Applicability Boundary
Eshkevari et al. [11]	Semantic mining and multi-criteria decision-making	Customer reviews	Constructs a computational chain for review ranking	Limited linkage between sentiment and satisfaction
Ganji et al. [12]	Deep learning and data balancing	Hotel reviews	Reduces the effect of class imbalance on classification training	Limited fine-grained semantic representation
Roy [13]	Sentiment analysis and lodging theory	Traveler online reviews	Links review sentiment with hotel performance	Weak rating prediction structure
Hidayati [14]	LDA and machine learning	Hotel reviews	Extracts aspect topics and identifies sentiment	Insufficient representation of contextual dependence
Marutho and Rustad [15]	Transformer and sparse attention	Aspect-level reviews	Strengthens sentence embedding and weight calculation	Relatively complex computational process
Jeong and Lee [16]	ChatGPT-based aspect-level analysis	Hotel service reviews	Supports semantic induction of service failures	Output constraints still need improvement
Malik and Bilal [17]	NLP survey and taxonomy framework	Online customer reviews	Summarizes the review processing workflow	Lacks validation with a single prediction model
Wankhade et al. [18]	Review of sentiment analysis methods	Multi-domain texts	Summarizes the technical evolution of sentiment analysis	Insufficient adaptation to tourism scenarios
Mustak et al. [19]	Machine-learning-based customer insight mining	User-generated content	Converts reviews into customer insights	Weak mapping of satisfaction levels
Hartmann et al. [20]	Sentiment analysis accuracy evaluation	Market review texts	Compares the application boundaries of sentiment models	Domain transfer remains limited

From the existing research, tourism review analysis has formed a complete basis for text processing. Machine learning models can complete sentiment classification, deep learning models can express context dependence, aspect-level analysis can identify service dimensions, and large language models have begun to participate in the semantic induction of reviews. Compared with the above studies, this paper focuses more on the joint modeling between sentiment polarity of tourism reviews and prediction of tourist satisfaction. The model does not separate sentiment classification, aspect recognition and rating prediction into isolated tasks, but completes review semantic encoding, sentiment feature extraction, multi-dimensional semantic fusion and satisfaction output in the same NLP deep network. The design can transform the review elements such as environment, service, price, traffic and facilities into weighted features, and use the shared semantic layer to maintain the consistency between emotional judgment and satisfaction level.

At the level of model design, tourism reviews also have the characteristics of short text, omitted expression and emotional transition. Both positive facility description and negative service feedback may occur in the same review, and a single word frequency vector can easily weaken the true judgment. Based on the literature, this paper introduces pre-training semantic coding, sequence dependence modeling and attention weight calculation, so that aspect words, sentiment words and adverbs of degree in reviews can enter the same feature space, and the output results are constrained by classification loss and regression loss. This processing method can reduce the interference of review topic drift on satisfaction prediction, and also preserve the implicit attitude formed by tone, negative words and comparative expressions in the text. The model output not only gives the sentiment category, but also converts the comment semantics into the satisfaction level, which provides the method basis for the multi-index verification in the subsequent performance analysis. At the same time, this process facilitates the migration and reuse between different tourism platforms, and maintains the stability of calculation results. The model constructed in this paper is closer to the real text processing scenario of online tourism platforms, and can provide a reusable computing framework for tourist experience recognition, intelligent review screening and service quality calculation.

3 Research Methods

3.1 Construction of tourist satisfaction prediction model based on NLP deep network

3.1.1 Tourism review text input layer and semantic preprocessing structure

The travel review text input layer is responsible for converting the original review into a sequence that can be read by the neural network. The sample covers scenic spots, hotels, transportation, catering and comprehensive service scenarios, and each record consists of the review body, star score, release time, service label and satisfaction level. The system removes repeated comments, advertising sentences and blank fields, and uniformly handles emotifications, abnormal punctuation, mixed capitalization and abbreviation expressions, so that the text maintains clear semantic boundaries. The length of the review is set to 128 words, and the short samples are filled with masks, and the long samples are retained according to the sentiment density.

Fig. 1 shows the input processing structure of travel review text before it enters the satisfaction prediction model. The original review was verified by fields, and empty text, repeated comments and abnormal rating records were deleted. Then it entered the noise filtering module to process advertising words, invalid symbols, garbled characters and continuous repeated expressions. Word segmentation, word form normalization and stop words filtering are performed on the cleaned text to form a computable word sequence. After the sequence length is unified, the system synchronously generates the word index matrix, location coding matrix and mask matrix, and sends them to the embedding mapping layer to provide a stable data entry for emotional semantic coding and tourist satisfaction prediction.

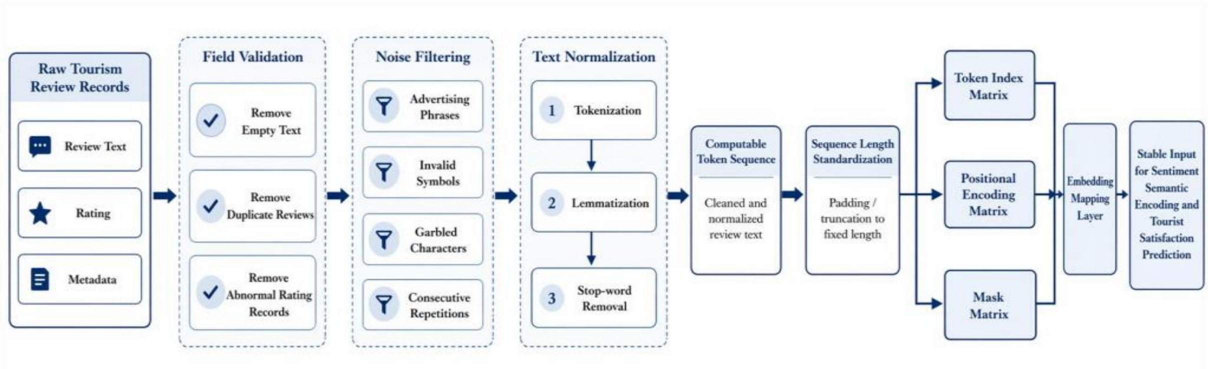


Figure 1: travel review text input layer with semantic preprocessing structure

In order to unify the input forms of reviews with different lengths, the input layer needs to write the words, positions and mask states into the sequence matrix at the same time to ensure the complete consistency of the dimension of batch training, which is specifically expressed as follows:

$$X_i = \{(w_{ij}, p_{ij}, m_{ij}) \mid j = 1,2, \dots, L\} \tag{1}$$

Here, X_i represents the input sequence of the i travel review, w_{ij} represents the j term, p_{ij} represents the position index of the term, m_{ij} represents the mask token, and L represents the uniform sequence length. This formula is used to fix the input form of comments, so that different lengths of texts enter the same batch training process, and reduce the interference of filled characters on the semantic encoding layer.

In view of the mixed situation of advertising words, garble code and low-frequency invalid words in tourism reviews, the preprocessing layer needs to establish the word retention function and noise constraint to stabilize the effective semantic boundary state. The specific calculation is as follows:

$$c_{ij} = \mathbb{I}\{\mathrm{freq}(w_{ij}) > \delta_1\} \cdot \mathbb{I}\{\mathrm{noise}(w_{ij}) < \delta_2\} \cdot \rho(w_{ij}) \tag{2}$$

Here, c_{ij} represents term retention coefficient, $\mathrm{freq}(w_{ij})$ represents term frequency statistic value, $\mathrm{noise}(w_{ij})$ represents noise intensity, δ_1 and δ_2 represent filtering threshold, $\rho(w_{ij})$ represents word form normalization function. The formula can delete invalid terms while retaining satisfaction related expressions such as service, price, environment and traffic, ensuring that the feature extraction has a clear object.

In order to make short sentences and long sentences comparable in the vectorization stage, the input layer needs to fuse word meaning, position and weight information to output semantic embedding. The specific process is as follows:

$$E_i = \sum_{j=1}^{L} \gamma_{ij} \left(\mathrm{Emb}(w_{ij}) + \mathrm{Pos}(p_{ij})\right), \quad \gamma_{ij} = \frac{\exp(q^{\top}\mathrm{Emb}(w_{ij}))}{\sum_{k=1}^{L} \exp(q^{\top}\mathrm{Emb}(w_{ik}))} \tag{3}$$

Here, E_i represents the semantic embedding vector output by the review input layer,

$Emb(w_{ij})$ represents the term embedding, $Pos(p_{ij})$ represents the position encoding, γ_{ij} represents the term weight, and q represents the learnable query vector. The formula aggregates word sense, word order and weight into a unified representation, which provides input for emotion encoding and satisfaction prediction. After processing, the input layer saves the term index matrix, mask matrix and semantic embedding matrix, and establishes the corresponding relationship with the emotion label and satisfaction label. Sentences containing negative words, transition words, and adverbs of degree retain adjacent positions, enabling the model to recognize mixed evaluations such as "good environment but too long queue".

After the above preprocessing, the tourism review text is transformed from loose natural language expression into model input with stable structure. The term index matrix preserves the review content, the position encoding matrix preserves the word order relationship, the mask matrix distinguishes the real terms from the filler terms, and the semantic embedding matrix undertakes the subsequent feature calculation of the deep network.

3.1.2 Emotional semantic encoding layer and context feature representation

The emotional semantic encoding layer is used to transform the term index matrix, the position encoding matrix and the mask matrix into the context semantic state. Tourism reviews often include service evaluation, environmental perception, price judgment, transportation experience and facility description at the same time, and it is difficult to fully express tourists' attitude with a single word. In this layer, the pre-trained language model is used as the underlying encoder, and the bidirectional sequence network and attention unit are superimposed, so that negation, turn, degree modification and aspect words are related in the same semantic space. After the input embedding enters the encoding layer, the system obtains the word-level vector and sentence-level representation, and outputs the feature matrix required for satisfaction prediction.

Fig. 2 shows the internal structure of the emotional semantic encoding layer. The word embedding and position encoding were firstly entered into the pre-trained semantic encoder to form the context word vector. Bi-directional context units compute left-right semantics so that turning expressions such as "good service but long queue" are not overridden by a single polarity. The attention layer assigns weights according to sentiment words, aspect words and degree adverbs, and the residual gating layer integrates the basic word meaning and deep sentiment representation.

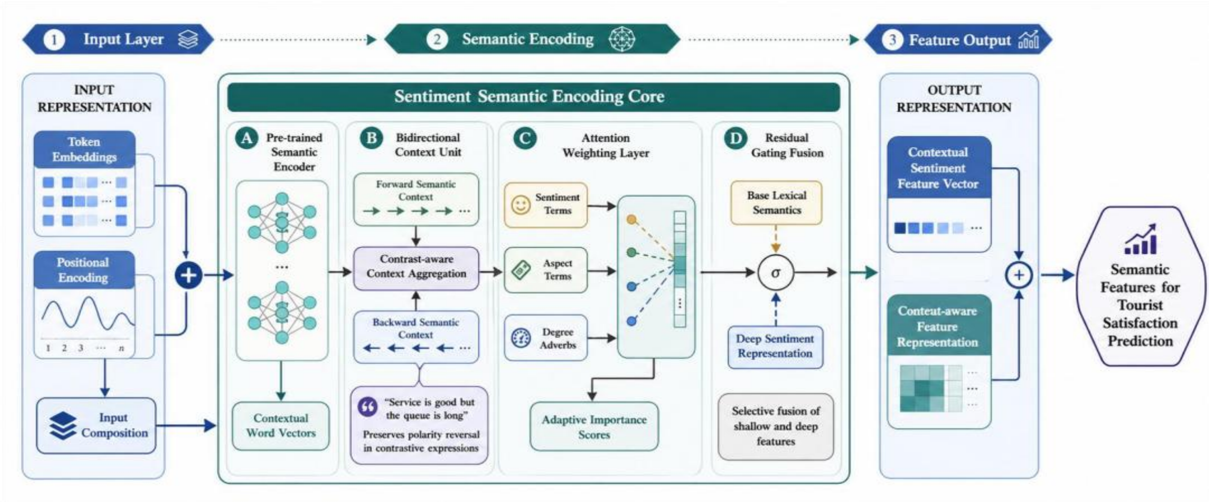


Figure 2: Emotion semantic encoding layer and context feature representation structure

In order to depict the emotional dependence of words in comments, the encoding layer adopts the form of bidirectional state update, retains the influence of word order change on attitude judgment, and forms the word-level context representation. The specific calculation is as follows:

$$\overrightarrow{h}_{ij} = f_\theta(e_{ij}, \overrightarrow{h}_{i,j-1}), \quad \overleftarrow{h}_{ij} = f_\theta(e_{ij}, \overleftarrow{h}_{i,j+1}), \quad h_{ij} = [\overrightarrow{h}_{ij}; \overleftarrow{h}_{ij}] \tag{4}$$

Here, e_{ij} represents the input embedding of the j word in the i comment, $\overleftarrow{h}_{ij}$ and $\overrightarrow{h}_{ij}$ represent the two types of semantic states, h_{ij} represents the concatenated context representation, and f_θ represents the sequence encoding function with parameters. This formula enables the model to read the semantic information before and after at the same time, and avoids one-way coding to weaken the emotional orientation in transition sentences and negative sentences.

In order to highlight the role of satisfication-related terms in the overall judgment, the attention layer generates normalized weights based on the semantic state of the terms and the sentiment query vector, and filters the core semantics, which is specifically expressed as follows:

$$a_{ij} = \frac{\exp((U_s h_{ij} + b_s)^\top r_s)}{\sum_{k=1}^{L} \exp((U_s h_{ik} + b_s)^\top r_s)}, \quad s_i = \sum_{j=1}^{L} a_{ij} h_{ij} \tag{5}$$

Here, a_{ij} represents the sentiment attention weight of the j term, U_s represents the semantic transformation matrix, b_s represents the bias term, r_s represents the sentiment query vector, and s_i represents the sentence-level sentiment semantic representation. The formula assigns different intensities to terms such as "poor, crowded, worthwhile, convenient", so that the prediction depends not only on the word frequency, but also on the semantic fragments that affect the experience judgment.

In order to alleviate the semantic drift caused by deep encoding, the model sets a residual gate structure at the output to maintain a controllable fusion state between the original embedding and the sentiment representation, which is calculated as follows:

$$g_i = \sigma(W_g[s_i; \bar{e}_i] + b_g), \quad z_i = g_i \odot s_i + (1 - g_i) \odot \bar{e}_i \quad \bar{e}_i = \frac{1}{L}\sum_{j=1}^{L} e_{ij} \tag{6}$$

Here, g_i represents the gating coefficient, W_g and b_g represent the learnable parameters, $\bar{e}_i$ represents the comment average embedding, z_i represents the context feature vector, and $\odot$ represents the element-wise multiplication. This formula establishes a regulatory relationship between deep emotional features and basic word senses, preserves the original semantics, and enhances the expression of satisfaction factors.

After encoding, the system writes z_i into the context feature matrix and enters the model training process together with the sentiment polarity label and the satisfaction level label. For multi-target reviews, the main sentiment segments are retained by attention weight. The basic semantic stability of short comments is maintained by residual gating. The processed attitude tendency, evaluation target and satisfaction cues can be expressed in a unified feature space, and the recognition of mixed emotion, implicit negation and degree change is more stable.

3.1.3 Tourist satisfaction prediction output layer and training target design

The satisfaction prediction output layer is used to convert the context feature vector into sentiment category, satisfaction level and continuous rating. This layer adopts a multi-task structure, and the sentiment classification branch and the satisfaction regression branch are set in parallel, so that the attitude tendency and the rating strength in the review are calculated on the same semantic representation. The model outputs three types of emotional probabilities and five levels of satisfaction prediction values, and integrates service, environment, price, traffic and facilities cues to form stable judgments and reliable output results.

In order to map contextual features into sentiment class probabilities, a normalized classification function is introduced into the output layer to estimate the probability of positive, neutral and negative sentiment and retain the class confidence difference, which is specifically expressed as follows:

$$\mathrm{p_i} = \mathrm{softmax}(\mathrm{W_c z_i} + \mathrm{b_c}), \quad \hat{\mathrm{y}}_\mathrm{i} = \arg\max_{\mathrm{r}\in\{1,2,3\}} \mathrm{p_{ir}} \tag{7}$$

Here, $\mathrm{p_i}$ represents the sentiment probability vector of the i comment, $\mathrm{W_c}$ represents the classification weight matrix, $\mathrm{b_c}$ represents the bias term, $\mathrm{b_c}$ represents the context feature vector, $\mathrm{p_{ir}}$ represents the r class sentiment prediction probability, and $\hat{\mathrm{y}}_\mathrm{i}$ represents the predicted sentiment label. The formula compresses the review semantics into a probability distribution, so that the model can distinguish the dominant sentiment and secondary sentiment, and avoid the mixture of reviews being simply classified into a single category.

In order to transform the comment semantics into continuous satisfaction scores, the regression branch uses a bounded mapping function to make the prediction results fall into a five-level evaluation interval stably and maintain the numerical interval between different semantic strengths. The specific calculation is shown in the following equation:

$$\hat{\mathrm{s}}_\mathrm{i} = 1 + 4\sigma(\mathrm{u_r^\top} \tanh(\mathrm{W_r z_i} + \mathrm{b_r}) + \beta_\mathrm{r}) \tag{8}$$

Here, $\hat{\mathrm{s}}_\mathrm{i}$ represents the prediction satisfaction score, $\mathrm{W_r}$ represents the regression mapping matrix, $\mathrm{u_r}$ represents the output weight vector, $\mathrm{b_r}$ and β_r represent the bias parameters, and σ represents the Sigmoid function. The formula maps the deep semantic features into one to five partitions, which can not only correspond to the scoring habits of the platform, but also avoid the output out of bounds caused by abnormal text.

To simultaneously constrain sentiment classification and satisfaction regression, a joint loss function is used as the training objective, which incorporates category cross-entropy, scoring error and parameter regularization term into the same optimization process and balance the gradient contribution of the two types of tasks, which is specifically expressed as follows:

$$\mathcal{L} = -\frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}}\sum_{\mathrm{r}=1}^{3} \mathrm{y_{ir}} \log \mathrm{p_{ir}} + \lambda_1 \frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}} \omega_\mathrm{i} (\mathrm{s_i} - \hat{\mathrm{s}}_\mathrm{i})^2 + \lambda_2 \|\Theta\|_2^2 \tag{9}$$

Here, $\mathcal{L}$ represents the joint training loss, N represents the total number of samples, $\mathrm{y_{ir}}$ represents the true sentiment label, $\mathrm{s_i}$ represents the true satisfaction score, ω_i represents the sample weights, Θ represents the set of model parameters, and λ_1 and λ_2 represent the loss weights. The formula enables the model to reduce the scoring bias when learning the sentiment boundary, and the sample weight can weaken the training bias caused by class imbalance.

In order to maintain the semantic consistency between sentiment polarity and satisfaction level, the consistency calibration term is set in the output layer, the sentiment probability is converted into a score correction factor, and the scores that deviate from the sentiment direction are smoothly corrected, as shown in the following equation:

$$\tilde{s}_i = \eta_i \hat{s}_i + (1-\eta_i)\left(3+\kappa\sum_{r=1}^{3} v_r p_{ir}\right), \quad \eta_i = \exp\left(-\left|\hat{s}_i - 3 - \kappa\sum_{r=1}^{3} v_r p_{ir}\right|\right) \tag{10}$$

Here, $\tilde{s}_i$ represents the satisfaction score after calibration, η_i represents the consistency coefficient, v_r represents the sentiment level mapping vector, and κ represents the correction strength. This formulation keeps negative reviews from being overestimated, positive reviews from being depressed, neutral reviews from being stationary, and controls the deviation between sentiment categories and satisfaction levels from the output.

After training, the output layer retains three types of results: sentiment probability, satisfaction score and consistency coefficient. The emotional probability explains the direction of tourists' attitude, the satisfaction score quantifies the experience intensity, and the consistency coefficient reflects whether the prediction results match the semantic tendency. The design integrates the review text, sentiment polarity and satisfaction score in the same target layer, so that the model can form a collaborative calculation relationship between classification and regression.

3.2 Sentiment feature extraction and polarity identification method of tourism reviews

The sentiment feature extraction and polarity identification method of tourism reviews is used to separate the evaluation target, sentiment word and intensity modification information from the encoded context vector. Tourism reviews often write service, environment, price, transportation and facilities in the same sentence, and the emotional direction is also affected by negative words and turning words. Firstly, the method locates the evaluation target according to the aspect dictionary and context representation, and then calculates the contribution degree of the term by using the sentiment query vector. Finally, it outputs three types of polarity probabilities: positive, neutral and negative, so that the review semantics can enter the satisfaction prediction link.

In order to form a computable association between aspect words and sentiment words in the same vector space, the model uses the semantic matching function to calculate the aspect-aware score, which is calculated as follows:

$$r_{im} = \frac{1}{L}\sum_{j=1}^{L}\tanh\left(h_{ij}^{\top} W_a a_m + b_a\right), \quad m = 1,2,\dots,M \tag{11}$$

Here, r_{im} represents the semantic matching strength of the i tourism review on the m evaluation aspect, $h_{ij}^{\top}$ represents the context status of the j word in the i review, a_m represents the vector representation of the m evaluation aspect, M represents the total number of evaluation aspects, W_a represents the aspect semantic mapping matrix, b_a represents the bias term, and L represents the uniform sequence length of the review. Through the matching calculation between the context state of the word and the aspect vector, the comment expressions such as "slow service", "convenient location", "high price" and "obsolete facilities" are mapped to specific evaluation dimensions, so that the subsequent sentiment polarity recognition has a clear semantic object.

In order to distinguish the influence of negation, adverbs of degree and turning connection on sentiment polarity, the model introduces modified gating coefficients and generates three types of sentiment logits, which are expressed as follows:

$$o_i = W_o \left[z_i; \sum_{m=1}^{M} r_{im}\, a_m; d_i; n_i; t_i \right] + b_o \tag{12}$$

Here, o_i represents the polarity logit vector, d_i represents the degree modification strength, n_i represents the negation marker strength, t_i represents the turning coefficient, W_o and b_o represent the classification parameters. The formula jointly calculates semantic states and linguistic modifiers, and can deal with mixed evaluations such as "good view but bad service".

To obtain a probabilistic output that can be used for classification, the polarity identification layer performs temperature scaling normalization on the logit and preserves the inter-class distance, which is calculated as follows:

$$P_i^c = \frac{\exp(o_i^c / \tau_p)}{\sum_{r=1}^{C} \exp(o_i^r / \tau_p)}, \quad c \in \{pos, neu, neg\} \tag{13}$$

Here, P_i^c denotes the c sentiment probability, τ_p denotes the temperature parameter, and C denotes the number of sentiment categories. Temperature scaling can reduce overconfidence and maintain reasonable probability distribution of boundary samples, which is especially suitable for neutral comments and short texts with unclear emotions.

In order to filter the low-confidence sentiment results, the model uses probability entropy and maximum class probability to form confidence constraints, and outputs the final polarity label, which is specifically expressed as follows:

$$q_i = \max_c P_i^c - \frac{-\sum_{c=1}^{C} P_i^c \log(P_i^c + \epsilon)}{\log C}, \quad \hat{c}_i = \arg\max_c P_i^c \tag{14}$$

where q_i represents the confidence, $\max_c P_i^c$ represents the maximum prediction probability in the sentiment class, $-\sum_{c=1}^{C} P_i^c \log(P_i^c + \epsilon)$ represents the entropy calculation term of the sentiment probability distribution, C represents the number of sentiment classes, ϵ represents the smoothing term to avoid logarithmic calculation anomalies, and $\hat{c}_i$ represents the final sentiment polarity label. This formula combines the maximum class probability with the normalized entropy value to distinguish the high confidence sentiment samples and semantic fuzzy samples.

After polarity identification, the model no longer only gives a single sentiment label, but retains the aspect strength, polarity probability and confidence results at the same time. For the comments with parallel evaluation such as "good environment but average service", the polarity recognition layer can distinguish the sentiment direction of different evaluation objects and avoid the whole review being simply classified as positive or negative. The confidence calculation also provides a basis for screening low certainty samples, so that noise reviews, semantic fuzzy reviews and emotional conflict reviews can be separately labeled, so that the emotional feature extraction results are closer to real tourists' expression.

3.3 Multi-dimensional semantic feature fusion mechanism for satisfaction prediction

The multi-dimensional semantic feature fusion mechanism for satisfaction prediction is used to integrate the evaluation target, sentiment polarity, semantic strength and rating prior in the review. The satisfaction judgment in tourism reviews is not only determined by a single emotion. Dimensions such as service, environment, price, transportation and facilities often appear at the same time, and may form mutually offsetting or mutually reinforcing relationships. The fusion layer needs to transform these decentralized semantics into structured features that enable the model to identify the actual contribution of different dimensions to the satisfaction result.

In order to establish the joint expression relationship between evaluation dimensions and sentiment polarity in tourism reviews, the fusion layer maps aspect strength, sentiment probability and context projection into a third-order semantic tensor, which is specifically expressed as follows:

$$\mathcal{T}_{i,m,c} = r_{im} \cdot P_i^c \cdot \psi_m(z_i), \quad m = 1,2,\dots,M, c = 1,2,\dots,C \tag{15}$$

Here, $\mathcal{T}_{i,m,c}$ represent the joint semantic tensor of the i review on the m evaluation dimension and the c sentiment class, P_i^c represents the sentiment class probability, and $\psi_m(z_i)$ represents the projection result of the context feature on the corresponding evaluation dimension. The formula can record composite semantics such as "service positive, price negative, traffic neutral", and avoid different evaluation dimensions being compressed into a single emotion value in the fusion stage.

In order to measure the actual contribution of different evaluation dimensions to satisfaction prediction, the fusion layer calculates the dimension contribution coefficient based on the strength distribution of the semantic tensor, and maintains the comparability between the dimensions, as shown in the following equation:

$$\pi_{im} = \frac{\sum_{c=1}^{C} \left|\mathcal{T}_{i,m,c}\right|}{\sum_{n=1}^{M} \sum_{c=1}^{C} \left|\mathcal{T}_{i,n,c}\right| + \varepsilon} \tag{16}$$

Here, π_{im} represents the contribution coefficient of the m evaluation dimension to the satisfaction prediction of the i review, and ε represents the smoothing term, which is used to avoid the denominator anomaly. The formula automatically assigns dimension contributions according to the content of reviews. When the service complaint is obvious, the weight of the service dimension will be enhanced, and when the price evaluation is prominent, the price dimension will occupy a higher contribution proportion.

In order to avoid semantic overlap of multiple evaluation dimensions in the fusion process, the model introduces dimension separation constraints to suppress the correlation of potential representations of different dimensions, as shown in the following equation:

$$\mathcal{D}_i = \sum_{m=1}^{M} \sum_{n=1 n\neq m}^{M} \frac{\left|u_{im}^{\top} u_{in}\right|}{\|u_{im}\|_2 \|u_{in}\|_2 + \varepsilon} \tag{17}$$

Here, $\mathcal{D}_i$ represents the dimension separation constraint term of the i review, u_{im} represents the latent semantic vector of the m evaluation dimension, and u_{in} represents the latent semantic vector of the n evaluation dimension. By inhibiting the excessive similarity between dimensions, the formula makes the factors such as service, environment, price, transportation and facilities remain independently expressed, and avoids the satisfaction prediction being overly dominated by a single dimension.

After the multi-dimensional semantic fusion is completed, the model does not obtain a single comment vector, but a composite feature representation including aspect contribution, sentiment distribution and dimension differences. This representation can preserve the relationship between different evaluation objects, and also reduce the interference of single emotion words on the overall judgment, so that the satisfaction prediction has a clear semantic source.

3.4 Tourist satisfaction prediction process based on review semantic weight

The tourist satisfaction prediction process based on review semantic weight is used to transform multi-dimensional fusion features into interpretable satisfaction results. Instead of repeating sentiment classification or simple score regression, the process revolves around semantic contribution recording, conflict calibration, and rank output. After reading the fusion features, the model forms a contribution trajectory for each evaluation dimension, checks whether the sentiment direction is consistent with the rating strength, and finally outputs a continuous score, a five-level level, and a traceable semantic basis.

Fig. 3 shows the tourist satisfaction prediction process. After fusing the features into the semantic contribution record layer, the system generated the contribution trajectories of services, environments, prices, transportation and facilities. The conflict calibration layer examines the deviation between sentiment polarity and rating changes. The level judgment layer maps the calibration results to the satisfaction level. The output retains the prediction score, rank label, dimension contribution, and confidence state.

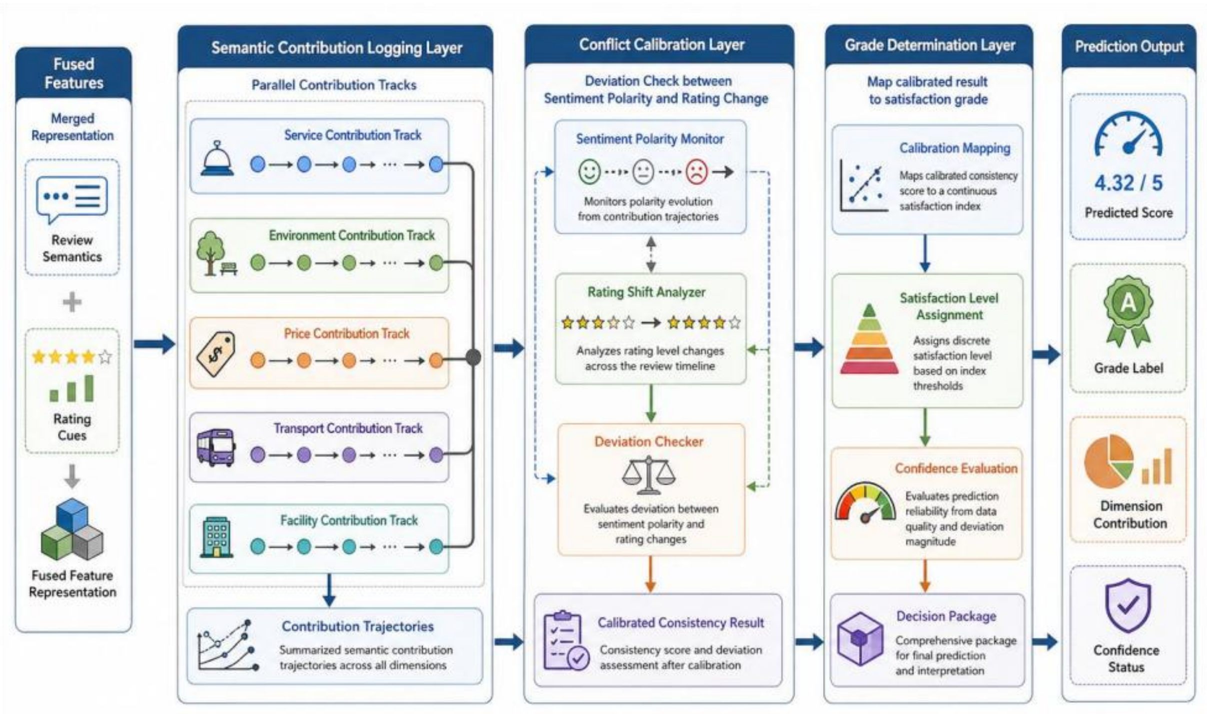


Figure 3: Visitor satisfaction prediction process based on review semantic weights

In order to record the contribution source of each evaluation dimension in the prediction process, the system organizes the dimension contribution coefficient, latent semantic vector and sentiment polarity direction into traceable records, which are specifically expressed as follows:

$$\mathcal{R}_i = \{(m, \pi_{im}, u_{im}, \varphi_{im})\}_{m=1}^{M}, \quad \varphi_{im} = \sum_{c=1}^{C} l_c p_i^c \pi_{im} \tag{18}$$

Here, $\mathcal{R}_i$ represents the semantic contribution record of the i comment, φ_{im} represents the sentiment contribution value of the m evaluation dimension, and l_c represents the sentiment category direction coefficient. The formula can save the model prediction basis, so that the satisfaction results correspond to specific evaluation dimensions, rather than only output unexplained values.

In order to deal with the samples with inconsistent sentiment direction and rating intensity, the process sets the conflict calibration coefficient to constrain the deviation between the semantic contribution and the initial score, which is calculated as follows:

$$B_i = \left| \mathrm{sgn}\left(\sum_{m=1}^{M} \varphi_{im}\right) - \mathrm{sgn}(s_i^{raw} - \bar{s}) \right| \cdot \frac{1}{M} \sum_{m=1}^{M} (1 - \pi_{im})^2 \tag{19}$$

Here, B_i represents the sentiment score conflict coefficient, s_i^{raw} represents the initial satisfaction score, $\bar{s}$ represents the median rating benchmark, and $\mathrm{sgn}(\cdot)$ represents the sign function. This formula is used to identify the samples with "significantly negative text but high score" or "positive text but low score", so that the prediction process can calibrate the abnormal samples.

In order to output stable five-level satisfaction results, the model maps the calibrated continuous scores into an ordered level space and retains the level boundary distance, which is specifically expressed as follows:

$$G_i = \arg\min_{k \in \{1,\dots,5\}} |\tilde{s}_i - \mu_k|, \quad b_i = \min_{k \neq G_i} |\tilde{s}_i - \mu_k| \tag{20}$$

Here, G_i represents the final satisfaction level,μ_k represents the μ_k satisfaction center value, and b_i represents the distance of the level boundary. This formula makes the continuous prediction results fall into the five-level evaluation system stably, and retains the judgment distance of critical samples.

After the tourist satisfaction prediction process is completed, the output of the system is not only the score and level, but also the semantic contribution record, the conflict calibration result and the level boundary distance. The continuous score is used to reflect the tourist experience intensity, the level label is used to correspond to the platform evaluation scale, the semantic contribution record is used to explain the prediction basis, and the boundary distance is used to determine whether the sample is close to the level critical interval. This process transforms tourism reviews from unstructured text to interpretable satisfaction prediction results, providing structured outputs for review ranking, service feedback screening, and visitor experience trend calculation.

4 Performance analysis

4.1 Sentiment classification of tourism reviews and analysis of satisfaction prediction results

The experiment uses 48620 English travel reviews as samples, and the sources cover scenic spots, hotels, transportation, catering and comprehensive service scenes. After cleaning, word form normalization, stop words filtering and rating label mapping, the data were divided into training set, validation set and test set according to 7：1：2. The model was run in Python3.10 and PyTorch environment, the input layer used 128 words length, the sentiment label was set to three categories of positive, neutral and negative, and the satisfaction label was mapped to 1-5 scores. This experiment examines the performance of NLP deep networks in sentiment classification and satisfaction prediction. AdamW optimizer was used for training with an initial learning rate of 2e-5, a batch size of 64 and a maximum number of training rounds of 30. The loss function consists of categorical cross-entropy and regression error, which enables the model to learn attitude polarity and rating strength.

Table 2 shows the three categories of emotion recognition results on the test set. The recognition accuracy of positive reviews reaches 95.31%, which indicates that the model has a strong ability to capture satisfactory expressions. The F1 value of negative comments is 92.44%, which has a stable discrimination ability for negative semantics such as poor service, high price and queuing time. Neutral reviews have a slightly lower indicator, mainly because such reviews contain weak sentiment and factual descriptions.

Table 2: Sentiment classification results for travel reviews

Sentiment Category	Number of Samples	Accuracy/%	Precision/%	Recall/%	F1-score/%
Positive	5124	95.31	94.86	95.02	94.94
Neutral	1876	91.42	90.18	89.67	89.92
Negative	2724	93.27	92.81	92.08	92.44
Overall	9724	94.18	93.47	92.86	93.15

Table 3 shows the satisfaction prediction results in different scenarios. The MAE of scenic spot reviews is 0.201, hotel reviews is 0.213, traffic reviews is 0.236, catering reviews is 0.219, and comprehensive service reviews is 0.224. The traffic scenario has a slightly higher error, which is related to time factors such as delay, transfer, and waiting. The overall MAE is 0.218, which is consistent with the summary results, indicating that the model has the ability to fit stable scores.

Table 3: Satisfaction prediction results for different tourism scenarios

Scenario Type	Test Samples	Actual Mean	Predicted Mean	MAE	RMSE
Scenic Spot Reviews	2360	4.18	4.15	0.201	0.286
Hotel Reviews	2145	4.06	4.02	0.213	0.301
Transportation Reviews	1688	3.74	3.69	0.236	0.337
Catering Reviews	1817	4.11	4.08	0.219	0.309
Comprehensive Service	1714	3.96	3.92	0.224	0.318
Overall	9724	4.02	3.98	0.218	0.310

Fig. 4 uses heat maps to present the semantic activation strength of evaluation dimensions in different emotion categories. In the positive reviews, service and environment had the highest heat, which were 0.82 and 0.79 respectively. In the negative comments, the heat of traffic and price were concentrated, which were 0.76 and 0.74, respectively. Neutral reviews were in the range of medium to low intensity, and none of the five dimensions exceeded 0.48.

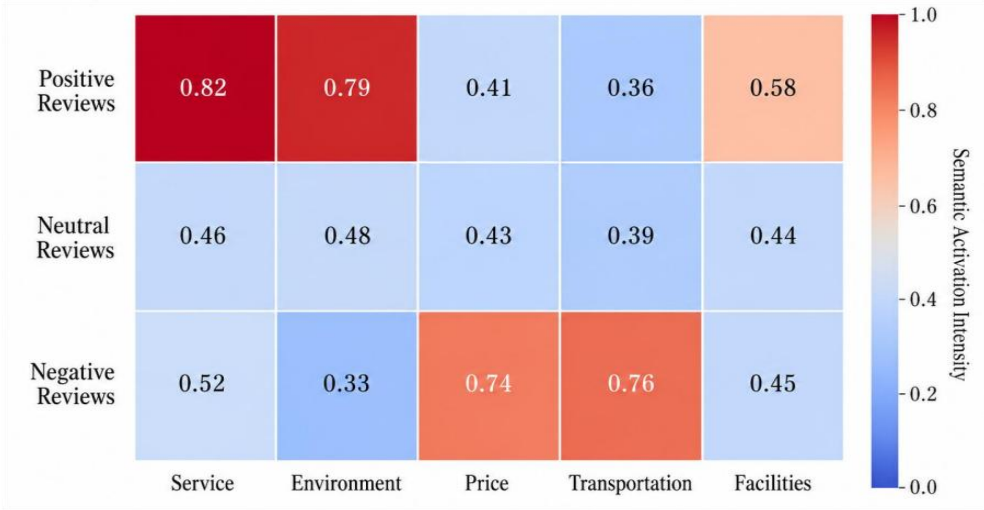


Figure 4: Heat map of evaluation dimension contribution under different sentiment categories

From the classification and prediction results, the proposed model achieves 94.18% sentiment classification accuracy on the test set, Precision is 93.47%, Recall is 92.86%, F1-score is 93.15%, and AUC is 0.962. The emotion recognition results are consistent with the satisfaction prediction results, which indicates that the semantic coding layer, polarity recognition layer and multi-dimensional fusion layer maintain cohesion. For compound comments such as "the view is good but the price is high", the model can retain the positive activation of the environment and negative activation of the price, and the prediction score will not be drawn by a single emotion word. This result illustrates that the joint modeling can handle mixed semantics and satisfaction judgments in travel reviews.

4.2 Multi-dimensional semantic evaluation analysis of tourist satisfaction

This section focuses on the evaluation of five semantic dimensions: service, environment, price, transportation and facilities. The model read the fusion features on 9724 travel reviews in the test set, and calculated the sentiment probability, aspect contribution and satisfaction score accordingly. At the same time, combined with the results of multiple rounds of training evaluation, the contribution changes of each semantic dimension in the process of model convergence were observed. The overall results show that the samples with high satisfaction scores are mostly supported by environment and service, and the samples with low satisfaction scores are mostly concentrated in semantic positions such as price burden, traffic waiting and insufficient service response. This analysis no longer stops at positive, negative or neutral labels, but investigates whether the model can transform the specific evaluation objects in tourist reviews into interpretable satisfaction bases.

To show the difference in the contribution of semantic dimensions under different emotion categories, Fig. 5 uses radar chart to present the average weight of the five dimensions. In the positive reviews, the weight of environment is 0.31, service is 0.27, facility is 0.16, price is 0.14, and transportation is 0.12. In the negative comments, the price increased to 0.29, the traffic was 0.26, the service was 0.24, the environment decreased to 0.13, and the facility was 0.08. The service, environment, price, transportation and facilities in the neutral comments are

0.21, 0.22, 0.19, 0.18 and 0.20, respectively, with a more balanced distribution, indicating that there is no obvious dominant dimension in the weak sentiment text.

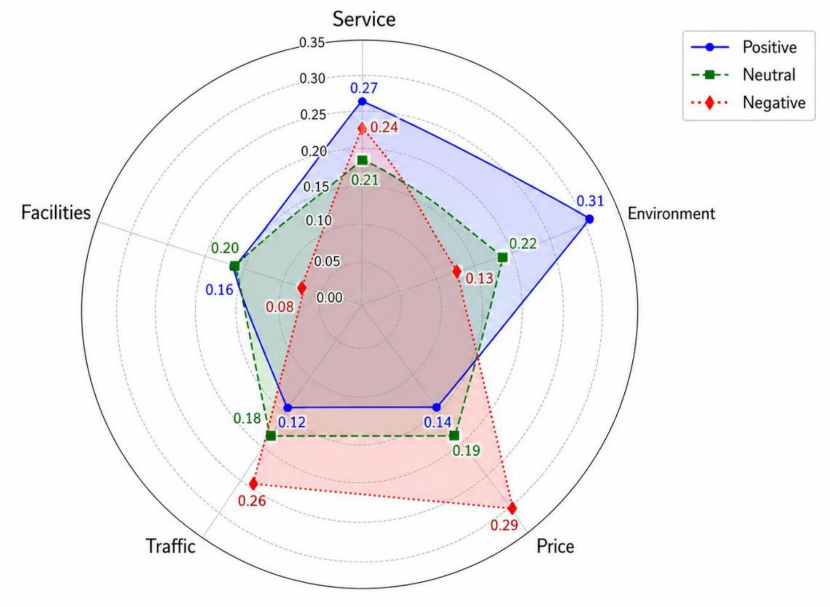


Figure 5: Radar plots of semantic dimensions of tourist satisfaction under different emotional categories

In order to observe the changes of semantic dimensions during training, Fig. 6 presents the five-round evaluation results using a multi-round comparison line chart. The service dimension increased from 0.62 to 0.78, and the environment dimension increased from 0.59 to 0.74, indicating that the model's recognition of service response, spatial experience and landscape perception was gradually stable. The facility dimension increased from 0.48 to 0.63, and the traffic dimension increased from 0.44 to 0.58, indicating that the semantics of supporting conditions, queuing and transfer were more clearly expressed in the later stage of training. The price dimension always has a negative contribution, falling from -0.42 to -0.61, indicating that expressions such as "expensive" and "worthless" will continue to drag down the satisfaction score. The overall trend shows that each semantic dimension gradually converges in multiple rounds of training, and the model can distinguish positive experience factors from negative cost factors.

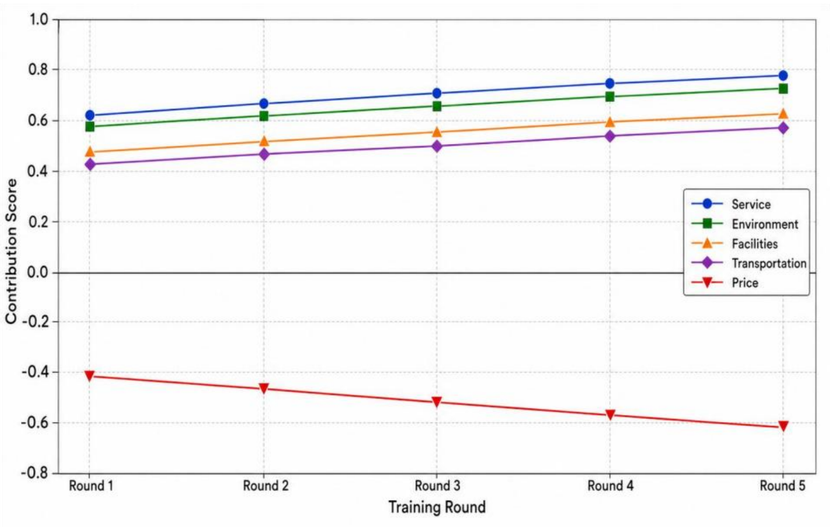


Figure 6: Line plot of contribution variation of different semantic dimensions in multiple rounds of training

The results of multi-dimensional semantic evaluation show that tourist satisfaction is not the direct output of a single emotional label, but is formed under the joint action of multiple evaluation dimensions. In the positive comments, spatial experience, service response and facilities are more likely to form high satisfaction judgments. In the negative comments, the traction of cost perception, waiting time and transportation convenience on the scoring results is more obvious. For the compound comments containing positive and negative juxtapositions, the model can retain the semantic differences of different dimensions and avoid a certain type of emotional words covering the overall judgment. After multi-dimensional semantic modeling, the satisfaction prediction has a clearer source of features, and can better correspond to the experience differences in real reviews.

4.3 Comparative analysis of tourism review categories based on semantic feature curves

In this section, the test set reviews are divided into five categories: scenic spots, hotels, transportation, catering and comprehensive services, and the semantic feature curves are used to observe the changes of different categories in sentiment polarity, satisfaction intensity and error distribution. The semantic feature curve is generated by the sentiment probability, aspect contribution coefficient and satisfaction prediction value, and the sample of each type of review is normalized after the average value. The analysis is used to test whether the model can distinguish the semantic structure in different tourism scenes, instead of only classifying positive and negative words.

In order to show the distribution differences of various types of comments in the semantic space, Fig. 7 uses a two-dimensional semantic trajectory density map to show the combined state of sentiment intensity and satisfaction residual. The comments of scenic spots mainly concentrated in the range of emotional intensity 0.72 -- 0.88 and residual error 0.12 -- 0.10, indicating that there were many samples with high emotional consistency. The trajectory of traffic comments spreads to the lower right region, and the lowest residual reaches -0.31, indicating that the waiting, delay and transfer semantics easily lower the score. The distribution of restaurant and hotel reviews is more concentrated, which indicates that service and price evaluation have a stable effect on satisfaction.

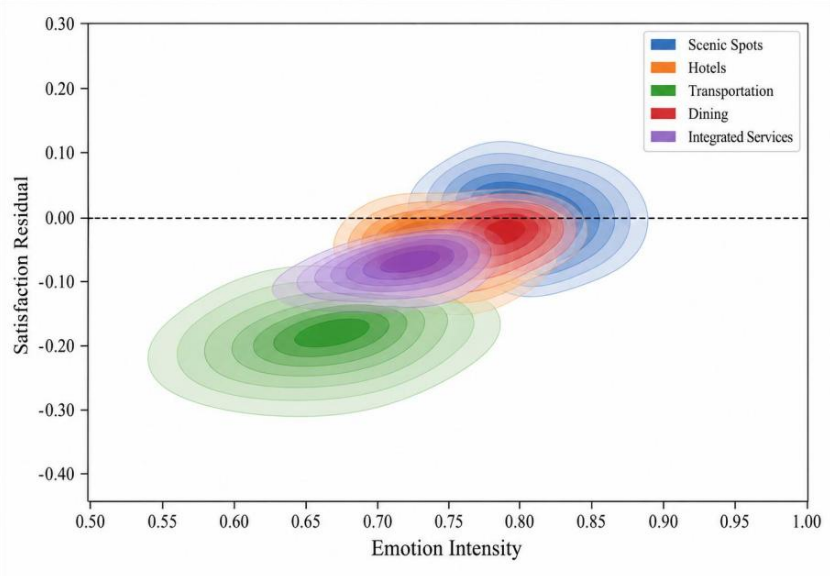


Figure 7: Density plot of semantic trajectories for tourism review categories

Table 4 further presents the statistical results of semantic curves for different categories. The peak value of scenic spot comments was 0.86, and the average satisfaction predicted value was 4.15. The peak value of traffic comments was 0.69, and the MAE reached 0.236. The curves for hotel and restaurant reviews are less volatile, with standard deviations of 0.118 and 0.121, respectively. The results show that the scene category affects the semantic concentration of reviews and also changes the error range of satisfaction prediction.

Table 4: Statistical results of semantic feature curves for tourism review categories

Review Category	Curve Peak	Semantic Standard Deviation	Predicted Mean	MAE
Scenic Spot Reviews	0.86	0.109	4.15	0.201
Hotel Reviews	0.81	0.118	4.02	0.213
Transportation Reviews	0.69	0.146	3.69	0.236
Catering Reviews	0.83	0.121	4.08	0.219
Comprehensive Service	0.76	0.132	3.92	0.224

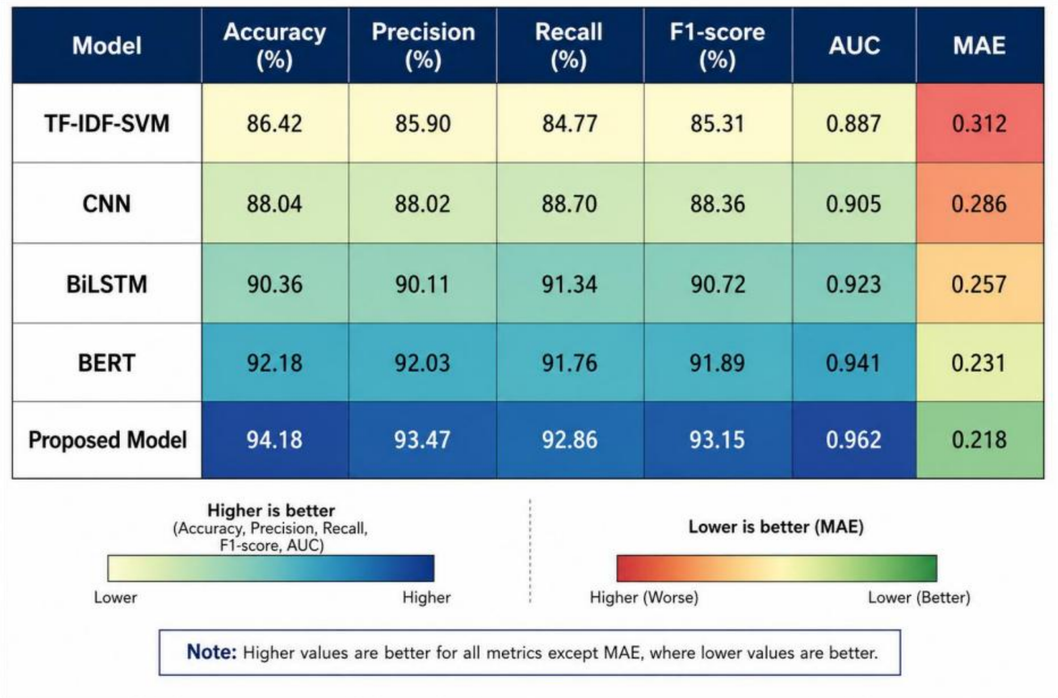
Further observation of the residual distribution of each category shows that the sample density of traffic reviews is the highest in the low satisfaction interval, and the negative words are mostly related to time cost. The comments of scenic spots are more concentrated in the high satisfaction interval, and the environment description and experience words keep changing in the same direction. The contributions of service words and facilities words in hotel reviews are similar, and the price words in catering reviews have a more obvious impact on the medium rating sample.

The results of semantic feature curve show that the model can transform the expression differences in different tourism categories into comparable computational features. Scenic reviews are more dependent on environment and experience words, traffic reviews are more affected by waiting costs, and hotel and catering reviews form a stable trade-off between service and price.

4.4 Comparative analysis of the proposed model and traditional text classification models

In this section, the proposed NLP deep network is compared with TF-IDF-SVM, CNN, BiLSTM, and BERT, and the evaluation content includes sentiment classification, satisfaction prediction, and model interpretable output. The comparison models use the same test set, and the text cleaning, label mapping and sample division are kept consistent. On the basis of pre-trained semantic coding, aspect recognition, multi-dimensional semantic fusion and consistency calibration are added to the model, so that sentiment probability and satisfaction score can be output in the same calculation process.

In order to compare the performance of different models in classification and prediction tasks, Fig. 8 presents the six indicators Accuracy, Precision, Recall, F1-score, AUC and MAE using the model performance matrix diagram. The Accuracy of TF-IDF-SVM is 86.42%, and the AUC is 0.887. The F1-score of CNN was 88.36%. The F1-score of BiLSTM is 90.72%. The Accuracy of BERT is 92.18%, and the AUC is 0.941. The Accuracy of the model in this paper reaches 94.18%, the F1-score is 93.15%, the AUC is 0.962, and the MAE is reduced to 0.218, indicating that it maintains a more stable comprehensive performance in sentiment classification and satisfaction prediction.



Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC	MAE
TF-IDF-SVM	86.42	85.90	84.77	85.31	0.887	0.312
CNN	88.04	88.02	88.70	88.36	0.905	0.286
BiLSTM	90.36	90.11	91.34	90.72	0.923	0.257
BERT	92.18	92.03	91.76	91.89	0.941	0.231
Proposed Model	94.18	93.47	92.86	93.15	0.962	0.218

Figure 8: Matrix plot of the performance of different text classification models

To examine the contribution of each module to the results, Table 5 presents the results of ablation experiments. After removing aspect recognition, the F1 value dropped to 91.02%, indicating that the evaluation object location would affect the polarity judgment. After removing the multi-dimensional semantic fusion, the MAE rises to 0.247, indicating that dimensions such as service, price and transportation cannot be simply merged. After removing the consistency calibration, the AUC drops to 0.948, and the scoring bias is more obvious in mixed sentiment reviews.

Table 5: Results of ablation experiments for the proposed model

Model Setting	Accuracy/%	F1-score/%	AUC	MAE
Without Aspect Recognition	91.36	91.02	0.934	0.239
Without Semantic Fusion	90.84	90.55	0.929	0.247
Without Consistency Calibration	92.47	91.88	0.948	0.231
Without Residual Gating	92.03	91.41	0.943	0.235
Full Model	94.18	93.15	0.962	0.218

The comparison results show that the traditional text classification model can complete the basic emotion recognition, but there are shortcomings in satisfaction prediction and composite semantic interpretation. Through the joint calculation of aspect-level semantics, dimension contribution and calibration mechanism, the proposed model makes comments such as "good view but long queue" and "slow service but reasonable price" get a more balanced rating mapping. The ablation results also prove that model performance is not contributed by a single encoder, but comes from the synergy between semantic recognition, feature fusion, and output calibration.

5 Discussion

The core of tourism review sentiment analysis and tourist satisfaction prediction is to convert scattered text into computable semantic evidence. The model retains negation, transition and degree modification in the preprocessing stage, and combines pre-training semantic representation, bidirectional context modeling and attention calculation in the encoding stage, so that the composite evaluation of "good scenery but long queue" and "high price but stable service" can be split into different semantic dimensions. Compared with TF-IDF-SVM, CNN, BiLSTM and BERT, the proposed model does not simply output sentiment categories, but synchronously forms aspect contributions, polarity probabilities, satisfaction scores and calibration results. In the experiment, the accuracy of the model reaches 94.18%, the F1 value reaches 93.15%, the AUC is 0.962, and the MAE is 0.218, which indicates that the semantic coding, dimension fusion and consistency calibration form a stable synergistic relationship. The multidimensional semantic analysis also shows that services and environments are more likely to support high satisfaction predictions, and price and traffic contribute more significantly in low-scoring samples. This result indicates that tourist satisfaction is not simply determined by positive and negative emotions, but is jointly shaped by specific evaluation objects, semantic intensity and rating direction. For online travel platforms, the model can be used for review screening, experience difference identification and service feedback ranking, and can retain the backtracking semantic basis to enhance the explanatory value of the prediction results. At the same time, the model deals with different scene reviews in a unified feature space, so that scenic spots, hotels, transportation and catering texts can be compared in the same evaluation framework, which is suitable for subsequent deployment into the large-scale tourism data analysis process.

6 Conclusions

This paper constructs an NLP deep network around tourism review sentiment analysis and tourist satisfaction prediction, and forms a computing link connecting text cleaning, semantic coding, polarity recognition, multi-dimensional fusion and satisfaction output. The experimental results show that the model can identify the evaluation dimensions of service, environment, price, transportation and facilities from the reviews of scenic spots, hotels, transportation, catering and comprehensive services, and convert them into emotional probability and satisfaction score. Compared with TF-IDF-SVM, CNN, BiLSTM, BERT and other models, the performance of the proposed method is more stable in composite reviews, weak emotional reviews, and rating bias samples, indicating that aspect-level semantic modeling, multi-dimensional feature fusion, and consistency calibration support the prediction results. This paper still has some limitations. The data mainly comes from English online travel reviews, and the language and platform sources are relatively concentrated. The adaptability of multi-language reviews, short video reviews and cross-cultural expression still needs to be verified. The model contains a pre-trained encoder and a multi-module fusion structure, and the inference cost is higher than that of the lightweight classifier. There is still room for compression of the response efficiency in real-time comment stream processing. The satisfaction label depends on the rating mapping, and there is a deviation between some review texts and user ratings. In the future, multi-language corpora, cross-platform reviews and time-series feedback data can be extended, knowledge distillation, pruning and parameter sharing methods are introduced to compress the model scale, and online learning mechanism is combined to update the semantic weights. Further, the review semantics can be associated with the click, stay, favorate and re-purchase behavior logs to test the applicability of the satisfaction prediction results in continuous behavior analysis and form the deployment application value.

Funding

This work was supported by the following grants:

1. Theoretical Logic and Practical Strategies of Digital Economy-driven High-quality Development of Rural Tourism in Anhui Province (Project No. 2025AHGXSK40183).
2. Research on the Construction of AI-Assisted Digital Multi-Dimensional Classroom under the OBE Concept: A Case Study of the Tourism Management Major (Project No. 2026jyyb03).
3. Mechanism Innovation and Realization Path of E-commerce Live Broadcasting Assisting the Integration of Agriculture, Culture and Tourism from the Perspective of Digital Economy (Project No. 2024AH053490).
4. Research on the Influence Mechanism of Live-streaming Marketing on the Image Construction of Tourism Destinations and Tourists' Travel Intentions (Project No. Wzyrc202407).

References

[1] Puh K, Bagić Babac M. Predicting sentiment and rating of tourist reviews using machine learning[J]. Journal of hospitality and tourism insights, 2023, 6(3): 1188-1204.

[2] Ameur A, Hamdi S, Ben Yahia S. Sentiment analysis for hotel reviews: a systematic literature review[J]. ACM Computing Surveys, 2023, 56(2): 1-38.

[3] Ali T, Omar B, Soulaimane K. Analyzing tourism reviews using an LDA topic-based sentiment analysis approach[J]. MethodsX, 2022, 9: 101894.

[4] Nawawi I, Ilmawan K F, Maarif M R, et al. Exploring tourist experience through online reviews using aspect-based sentiment analysis with zero-shot learning for hospitality service enhancement[J]. Information, 2024, 15(8): 499.

[5] Zakarija I, Škopljanac-Mačina F, Marušić H, et al. A sentiment analysis model based on user experiences of Dubrovnik on the Tripadvisor platform[J]. Applied Sciences, 2024, 14(18): 8304.

[6] Hanić S, Bagić Babac M, Gledec G, et al. Comparing machine learning models for sentiment analysis and rating prediction of vegan and vegetarian restaurant reviews[J]. Computers, 2024, 13(10): 248.

[7] Charfaoui K, Mussard S. Sentiment analysis for tourism insights: A machine learning approach[J]. Stats, 2024, 7(4): 90.

[8] George O A, Ramos C M Q. Sentiment analysis applied to tourism: exploring tourist-generated content in the case of a wellness tourism destination[J]. International journal of spa and wellness, 2024, 7(2): 139-161.

[9] Díaz-Pacheco Á, Guerrero-Rodríguez R, Alvarez-Carmona M A, et al. A comprehensive deep learning approach for topic discovering and sentiment analysis of textual information in tourism[J]. Journal of King Saud University-Computer and Information Sciences, 2023, 35(9): 101746.

[10] Collini E, Nesi P, Pantaleo G. Reputation assessment and visitor arrival forecasts for data driven tourism attractions assessment[J]. Online Social Networks and Media, 2023, 37: 100274.

[11] Eshkevari M, Rezaee M J, Saberi M, et al. An end-to-end ranking system based on customers reviews: Integrating semantic mining and MCDM techniques[J]. Expert Systems with Applications, 2022, 209: 118294.

[12] Ganji R N, Dadkhah C, Tohidi N. Improving sentiment classification for hotel recommender system through deep learning and data balancing[J]. Computación y Sistemas, 2023, 27(3): 811-825.

[13] Roy G. Travelers' online review on hotel performance–analyzing facts with the theory of lodging and sentiment analysis[J]. International Journal of Hospitality Management, 2023, 111: 103459.

[14] Hidayati N N. Improving aspect-based sentiment analysis for hotel reviews with Latent Dirichlet Allocation and machine learning algorithms[J]. Register: Jurnal Ilmiah Teknologi Sistem Informasi, 2023, 9(2): 144-159.

[15] Marutho D, Rustad S. Optimizing aspect-based sentiment analysis using sentence embedding transformer, bayesian search clustering, and sparse attention mechanism[J]. Journal of Open Innovation: Technology, Market, and Complexity, 2024, 10(1): 100211.

[16] Jeong N, Lee J. An aspect-based review analysis using ChatGPT for the exploration of hotel service failures[J]. Sustainability, 2024, 16(4): 1640.

[17] Malik N, Bilal M. Natural language processing for analyzing online customer reviews: a survey, taxonomy, and open research challenges[J]. PeerJ Computer Science, 2024, 10: e2203.

[18] Wankhade M, Rao A C S, Kulkarni C. A survey on sentiment analysis methods, applications, and challenges[J]. Artificial intelligence review, 2022, 55(7): 5731-5780.

[19] Mustak M, Hallikainen H, Laukkanen T, et al. Using machine learning to develop customer insights from user-generated content[J]. Journal of Retailing and Consumer Services, 2024, 81: 104034.

[20] Hartmann J, Heitmann M, Siebert C, et al. More than a feeling: Accuracy and application of sentiment analysis[J]. International Journal of Research in Marketing, 2023, 40(1): 75-87.