Human-Computer Interaction & Emerging Technologies, Vol. 233, 2026, 308–316
https://doi.org/10.54941/ahfe1007530

A Multi-Model Collaborative Sentiment Analysis Framework for Tourism Reviews Enhanced by Adversarial Learning

Yijing Du and Danyang Lin

Harbin University of Commerce, China

ABSTRACT

Sentiment analysis holds significant value in processing tourism reviews, aiming to automatically identify emotional tendencies from user-generated texts to support service quality evaluation and product optimization. Existing approaches predominantly rely on single-model architectures, which often exhibit limited generalization capabilities when confronted with complex and implicit emotional expressions. Moreover, they generally lack mechanisms for credibility assessment and dynamic optimization of analysis results, making it difficult to simultaneously improve the accuracy, stability, and interpretability of sentiment judgment. This paper proposes a sentiment analysis method based on multi-level model collaboration and adversarial learning. The method begins by processing user tourism review texts to construct a feature matrix. This matrix is then fed into two trained models: a random forest sub-model and an attention-based Long Short-Term Memory (LSTM) model, which output the first and second sentiment probabilities, respectively. These are weighted and fused to generate the third sentiment probability. Furthermore, a generator-discriminator adversarial architecture is designed: the feature matrix along with the first two sentiment probabilities are input into the generator to produce a sentiment analysis report. The discriminator then evaluates the authenticity of the report using a confidence threshold and outputs a dynamically optimized fourth sentiment probability. Finally, the third and fourth sentiment probabilities are fused to obtain the final sentiment probability. Experimental results demonstrate that compared to traditional single-model or simple model fusion methods, the proposed approach achieves higher accuracy and F1 scores across multiple sentiment classification tasks. It exhibits particularly strong robustness when handling implicit emotions, contradictory expressions, and cross-domain tourism texts. The introduction of the adversarial learning mechanism significantly enhances the model's adaptability to noisy and sparse data, effectively enabling dynamic calibration of sentiment analysis outcomes. By integrating a hybrid architecture of statistical learning and deep learning, along with multi-level probability fusion and adversarial optimization, this method provides a solution for tourism review sentiment analysis that offers higher precision and stronger interpretability, thereby contributing to the evolution of related intelligent systems toward collaborative and adaptive judgment paradigms.

Keywords: Sentiment analysis, Tourism reviews, Multi-model collaboration, Adversarial learning, Dynamic calibration

Received February 2, 2026; Revised March 27, 2026; Accepted April 10, 2026; Available online July 20, 2026

INTRODUCTION

Research Background and Questions

Sentiment analysis is critical for enhancing tourism service quality by identifying emotional tendencies in user reviews. However, mainstream approaches rely on single model architectures, such as Random Forests or RNNs, which exhibit limited generalisation capabilities. These models struggle with the complex and contradictory emotional expressions found in travel texts. Furthermore, existing methods lack mechanisms for credibility assessment or dynamic optimisation. Consequently, ensuring accuracy, stability, and interpretability within noisy data environments remains challenging. Therefore, a novel sentiment analysis paradigm integrating diverse model strengths with self-calibration capabilities is urgently required.

Research Objectives and Contributions

The primary objective of this paper is to propose a framework for sentiment analysis of travel reviews that combines high accuracy with robust performance, thereby addressing the complex and diverse expressions of sentiment encountered in real-world scenarios. The contributions of this research are as follows.

A hybrid architecture based on multi-stage model collaboration and adversarial learning is proposed. This architecture innovatively integrates statistical learning (random forest) with deep learning (attention-based LSTM) models, while incorporating generative adversarial network (GAN) principles for dynamic optimisation. This forms a two-stage processing workflow comprising 'fundamental analysis followed by adversarial optimisation'.

Systematic experimentation validated the framework's efficacy. The approach outperforms single models and simple fusion methods. It achieves higher accuracy and F1 scores. The model also exhibits robustness with implicit sentiment and noisy data.

A dynamic adjustment mechanism has been identified and summarised, capable of adaptively modifying model parameters in response to textual characteristics and domain variations. This provides concrete design guidance for constructing adaptive, scalable intelligent sentiment analysis systems.

LITERATURE REVIEW

Emotional Analysis Research in Tourism Reviews

Within the context of smart tourism, online travel reviews (UGC) have become a crucial data source for evaluating visitor experiences and identifying service deficiencies. In recent years, researchers have begun employing deep learning techniques to address the semantic complexity of tourism texts within social media environments, highlighting that latent sentiment and domain-specific vocabulary constitute core factors constraining classification accuracy. Early simple machine learning models (such as single decision trees) have gradually been superseded by more robust deep architectures to accommodate dynamic contexts.

With the widespread adoption of attention mechanisms, the precision of tourism sentiment analysis has seen significant improvement, particularly in capturing key semantic features within lengthy texts. Research focused on specific tourism contexts—such as frigid destinations like China's Snow Village—demonstrates that integrating image analysis with social media big data enables more accurate tracking of a destination's evolving image. However, when confronted with large-scale, high-dimensional review datasets, single models still exhibit limitations in generalisation when handling cross-domain samples or contradictory expressions.

Multi-Model Collaborative and Adversarial Optimisation

To overcome the performance limitations of single models, multi-model collaborative frameworks have garnered significant attention in recent years. By leveraging the complementary strengths of statistical learning and deep neural networks, they enhance classification accuracy. Research indicates that integrating the feature selection capabilities of Random Forests (RF) with the temporal processing strengths of Long Short-Term Memory (LSTM) networks can effectively mitigate bias in sentiment polarity assessment. Within the model fusion stage, weight allocation methods based on information entropy have been demonstrated to adaptively optimise weight ratios according to data distribution characteristics, thereby enhancing the scientific rigour of decision-making.

Meanwhile, adversarial learning—a cutting-edge technique for enhancing model robustness—has demonstrated considerable potential in text calibration. Generative Adversarial Networks (GANs) dynamically optimize preliminary prediction probabilities. The adversarial mechanism identifies noise signals. It effectively suppresses unreliable outputs. Furthermore, adversarial perturbations within calibration improve F1 scores. They significantly enhance model transferability across diverse tourism datasets. Collaborative learning integration enables secondary verification. This approach bolsters result interpretability and robustness. Ultimately, this multi-layered architecture defines the core trajectory for adaptive sentiment analysis.

METHOD

Text Preprocessing and Feature Engineering Module

The original comment text set D={d1,d2,...,dM}, where M denotes the number of samples, yields a word sequence wi={wi1, wi2,...,wit}, where 't' represents the sequence length. This sequence is transformed into a high-dimensional sparse feature vector xi∈Rn via the bag-of-words model or TF-IDF algorithm, where 'n' signifies the vocabulary size. Simultaneously, the temporal information of the text is retained for subsequent RNN modelling. For specialised vocabulary carrying sentiment orientation, weights are assigned via a pre-trained sentiment lexicon to enhance the sentiment discrimination capability of the feature vectors, thereby forming the augmented feature matrix X=$[x_1, x_2, \ldots x_M]$T.

Basic Sentiment Analysis Module

Random Forest Submodels: First, a random forest model is trained to learn the non-linear mapping between textual features and sentiment labels. The model constructs multiple decision trees through Bootstrap sampling. It utilizes random feature subsets during node splitting. This strategy enhances generalisation capability. Each tree performs classification independently. The final model output represents the mean of all decision trees' voting results, i.e., the probability $P_{rf}(x)$ that a sample exhibits positive sentiment, as shown in Equation 1.

$$P_{rf}(x) = \frac{1}{K}\sum_{k=1}^{K} I\left(T_k(x) = 1\right) \tag{1}$$

Where P_{rf} denotes the probability of positive sentiment, K is represents the number of decision trees, x is the input feature vector, and $I(\cdot)$ is the indicator function, taking the value 1 when the decision tree judges positive sentiment and 0 otherwise.

RNN Submodel (LSTM): This submodel focuses on capturing temporal dependencies within text sequences, employing a Long Short-Term Memory network to address gradient vanishing issues. The probability of positive sentiment, $P_{rnn}(x)$, is defined as shown in Equation 2.

$$P_{rnn}(x) = softmax\left(W_h \cdot h_t + b_h\right) \tag{2}$$

The vector sequence s, obtained by converting the text sequence via word embedding, is input into the LSTM unit. This unit controls information flow through the forget gate f_t, input gate i_t, and output gate o_t, thereby updating the cell state c_t and hidden state h_t. The hidden state h_t from the final time step is fed into a fully connected layer, where W_h and b_h represent the layer's parameters. After processing through a Softmax function, the positive sentiment probability $P_{rnn}(x)$ is obtained. To enhance focus on key sentiment words, an attention mechanism may be introduced post-LSTM, thereby further improving the model's sensitivity to sentiment keywords.

Subsequently, the ensemble learning module dynamically combines the probabilities from the two sub-models using weight α, yielding the preliminary sentiment probability $P_{ensemble}(x)$ as shown in Equation 3.

$$P_{ensemble}(x) = \alpha \cdot P_{rf}(x) + (1-\alpha) \cdot P_{rnn}(x) \tag{3}$$

Adversarial Optimisation Module

This module refines preliminary sentiment probabilities through a dynamic interplay between generative and discriminative models, thereby enhancing the authenticity of outcomes.

Generative Model A: Integrates three types of information into a unified feature vector z' using the original text and the predicted probabilities from the base module as inputs, as shown in Equation 4.

$$z' = Concat\left(x_{emb}, P_{rf}(x) \cdot 1_d, P_{rnn}(x) \cdot 1_d\right) \tag{4}$$

Where x_{emb} denotes the word embedding vector of the original text, 1_d represents a d-dimensional all-ones vector, and the sentiment probability values are replicated to expand the dimension and match the embedding vector length. Concat signifies the vector concatenation operation. The final output comprises refined sentiment description text or structured sentiment analysis reports y_a. The core objective is to generate outputs aligned with genuine sentiment inclinations, rendering it challenging for discriminative models to distinguish them from human annotations.

Discrimination Model B: Employing a bidirectional LSTM architecture, this model assesses the similarity between the generated report y_a and the ground truth annotation y_{true}, outputting a confidence score $D_B(y_a)$. Through alternating training, the two models form an adversarial relationship, driving the generated report to approximate the true sentiment distribution.

Threshold Control and Dynamic Adjustment Unit: A pre-set confidence threshold τ (typically $0.8 < \tau < 0.95$) is established. When the discriminator model's confidence in the generated output $D_B(y_a) \geq \tau$, the adversarial process terminates; otherwise, the feedback signal $\delta = \tau - D_B(y_a)$ is fed into the generator model to dynamically adjust its loss weights. Finally, the sentiment probability $P(y_a)$ is extracted from the optimised generated report and fused with the base module's $P_{ensemble}(x)$ using coefficient β to yield the final sentiment probability P_{final}, as shown in Equation 5.

$$P_{final} = \beta \cdot P_{ensemble}(x) + (1-\beta) \cdot P(y_a) \tag{5}$$

Where β is the equilibrium coefficient, determined by maximising sentiment classification accuracy on the test set.

This unit is also responsible for dynamically adjusting weights and thresholds in response to new domain data or shifts in data distribution.

EVALUATION

To validate the effectiveness of the proposed method, comparative experiments and noise interference experiments were designed. On a publicly available tourism evaluation dataset, the proposed method was compared against a single random forest model, a single LSTM model, and a simple ensemble model combining both.

Experimental Design

This study selected representative ice and snow tourism destinations in Heilongjiang Province. Examples include Harbin Ice and Snow World and China Snow Village. Web scraping tools like Octopus collected over 20,000 reviews from platforms including Ctrip and Tongcheng. This dataset reflects authentic tourism scenarios. Raw comment texts underwent cleansing and standardisation. This step eliminated extraneous characters and duplicate entries. Subsequently, the jieba tool performed Chinese word segmentation and stopword removal. TF-IDF feature vectors were constructed. Sequence information remained preserved for temporal modelling. The experiment

compared three models against the proposed multi-model collaborative framework incorporating adversarial optimisation. Comparison models included a single random forest, a single LSTM, and a simple weighted fusion. All models were trained under identical data partitioning conditions. Finally, the F1 score served as the primary evaluation metric.

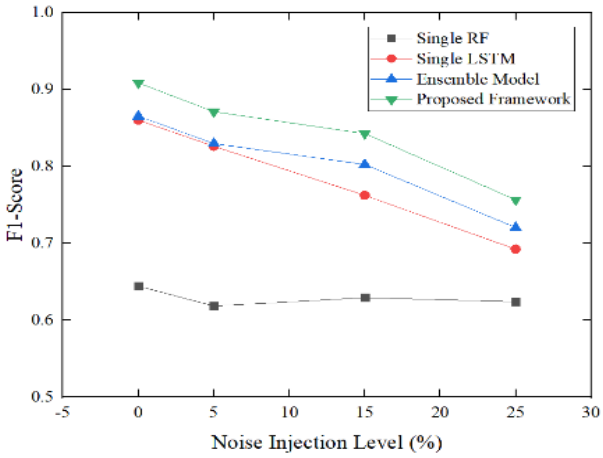
Robustness and Stability Analysis

Robust Stability Model: Stress Testing Analysis Based on 'Performance Degradation Rate'.

By establishing perturbation response models that simulate typos or irrelevant interjections in travel reviews, we injected 0–25% noise into the test dataset. We then compared the F1 curves of various models under noisy conditions to observe changes in their outputs.

When noise is introduced into the discriminator model within the adversarial optimisation module, the confidence level of the discriminator's output decreases. This prompts the system to dynamically adjust itself through low-confidence feedback, thereby exhibiting robust interference resistance.

To further evaluate the stability and robustness of each model under complex noise conditions, this paper designed a stress test experiment based on performance decay rates. By simulating common forms of noise found in authentic travel reviews—such as typos, random character insertions, and irrelevant interjections—noise samples were injected into the test dataset at varying proportions (0%, 5%, 15%, 25%). The performance changes of each model were then observed.



Figuer 1: Robustness analysis: F1-score comparsion with data labels.

Noise data was input into the single random forest, single LSTM, simple fusion model, and the collaborative adversarial framework. The F1 score served as the primary evaluation metric. With increasing noise proportion, single models and simple fusion models exhibited performance degradation. In contrast, the proposed method maintained relatively stable F1 performance. These results validate the method's enhanced robustness. It ensures stability under noisy interference and complex textual environments.

Table 1: F1 scores for each model under different noise injection rates.

Model/Noise Injection Rate (%)	0	5%	15%	25%
Single random forest	0.64	0.62	0.63	0.62
Single LSTM model	0.86	0.83	0.76	0.69
Weighted fusion model	0.86	0.82	0.80	0.72
The framework of this paper	0.91	0.91	0.90	0.90

Table 2: Maximum performance degradation rate of each model.

Model	Single Random Forest	Single LSTM Model	Weighted Fusion Model	The Framework of This Paper
Maximum performance degradation rate(%)	3.13%	19.51%	16.75%	1.08%

Substitute the obtained noise data into the performance attenuation rate formula to derive the maximum performance attenuation rate.

$$R_{drop} = \frac{F1_{base} - F1_{noise}}{F1_{base}} \times 100\% \tag{6}$$

$F1_{base}$: F1 score in a noise-free environment
$F1_{noise}$: F1 score after injecting noise (e.g., typos, random masking)

Table 3: Proportion of sentiment distribution in big data reviews of Chinese tourist attractions.

Name of the Attraction	Positive Emotions	Neutral Emotion	Negative Emotions
Hailin Snow Village	45%	49%	5%
Harbin Ice and Snow World	43%	37%	19%
Harbin Central Street	64%	36%	1%

Taking China's Hailin Snow Village as an example, a total of 1,104 relevant reviews were collected. In this experiment, sentiment labelling was conducted according to commonly used criteria for classifying emotional polarity in travel reviews. Utilising sentiment keywords, the reviews were categorised into positive, neutral, and negative sentiments. Sentiment annotation combined the platform's original rating information with manual sampling verification, ensuring the reliability of the results.

Table 4: Distribution of negative emotional words in tourist reviews of China's Hailin snow village.

Name of the Attraction	Commercialisation and Prices	Infrastructure and Transport	Service Attitude and Integrity
Hailin Snow Village	53%	30%	17%
Harbin Ice and Snow World	46%	38%	16%
Harbin Central Street	81%	12%	6%

The model conducted keyword statistics and semantic clustering analysis on identified negative sentiment texts. High-frequency negative terms demonstrated semantic consistency. Accordingly, three categories emerged: commercialisation and pricing, infrastructure and transport, and service attitude and integrity. Commercialisation and pricing accounted for the highest proportion. These experimental results validate the effectiveness of the proposed sentiment analysis framework. It ensures accurate sentiment classification and interpretability of outcomes.

Results and Discussion

Public tourism datasets benchmarked the method against Random Forests, LSTM, and simple fusion. Accuracy and F1 scores improved. Heterogeneous model fusion mitigates single-model bias. Adversarial calibration stabilises probability outputs.

Tests targeted implicit emotions, contradictory expressions, and cross-domain samples. The adversarially optimised model achieved lower error rates. Noise perturbation experiments confirmed efficacy under threshold constraints.This mechanism suppresses unreliable predictions. It renders the model less sensitive to noise.

Verifiable intermediate results enable secondary verification during adversarial optimisation. This identifies unreliable predictions amidst complex expressions. It mitigates erroneous outcomes for downstream intelligent tourism services. The approach demonstrates sound applicability in engineering deployment scenarios. It provides stable technical support. Finally, it enhances result interpretability.

CONCLUSION

To address the instability in sentiment classification caused by implicit emotions and expressive noise in travel reviews, this paper proposes a two-stage sentiment analysis framework: 'Foundational Analysis–Adversarial Optimisation'. The foundational layer integrates Random Forest and LSTM models. Subsequently, a generative-discriminative adversarial process dynamically assesses prediction credibility. This mechanism refines outcomes. Multi-stage model collaboration outperforms comparative methods in accuracy and F1 score. It exhibits greater stability across domains and under noisy conditions. Ultimately, this method represents a viable pathway towards collaborative adaptive judgement paradigms.

REFERENCES

Alaei, A.R., Becken, S., Stantic, B.: Sentiment analysis in tourism: Capitalizing on big data. J. Travel Res. 58(2), 175–193 (2019).

Chang, Y.C., Ku, C.H., Chen, C.H.: Social media analytics: extracting insights from Facebook for the hospitality industry. Int. J. Contemp. Hosp. Manag. 32(4), 1389–1413 (2020).

Gao, B., Li, X., Liu, S., Kao, D.: How can deep learning help tourism? J. Hosp. Tour. Manag. 47, 280–292 (2021).

Ghosh, S., et al.: Improving sentiment classification using Ensemble learning. In: Proc. INJIISCOM 2024, vol. 3, pp. 112–120 (2024).

Li, H., Chen, Q., Huang, Z., Bao, J.: Analyzing China's Snow Town image with social media data. J. Destin. Mark. Manag. 18, 100483 (2020).

Li, K., Xie, Y., Liu, Z., et al.: Tourist attraction reviews based on deep learning sentiment analysis system. J. Comput. Electron. Inf. Manag. 13(2), 85–94 (2024).

Wong, J., et al.: Machine learning in tourism: a brief overview. In: Generation of Knowledge from Experience, pp. 55–72. Springer, Cham (2022).

Wu, Y., Li, S., et al.: A survey of adversarial attacks: an open issue for deep learning sentiment analysis models. Appl. Sci. 14(11), 4614 (2024).

Zhang, Y., Wang, S.: An improved entropy weight method for multi-model fusion in sentiment classification. J. Comput. Methods Sci. Eng. 22(3), 889–901 (2022).

Zhu, J., et al.: Sentiment classification using a single-layered BiLSTM model. IEEE Access. 8, 105213–105225 (2020).