International Journal of Information Technology and Web Engineering
Volume 18 • Issue 1

A TBGAV-Based Image-Text Multimodal Sentiment Analysis Method for Tourism Reviews

Ke Zhang, College of Economics and Management, Neijiang Normal University, Neijiang, Sichuan, 641100, China & Faculty of Business, Economics and Accountancy, Universiti Malaysia Sabah, Kota Kinabalu, Sabah, 88400, Malaysia*

Shunmin Wang, College of Economics and Management, Neijiang Normal University, Neijiang, Sichuan, 641100, China

Yuanyu Yu, College of Artificial Intelligence, Neijiang Normal University, Neijiang, Sichuan, 641100, China

ABSTRACT

To overcome limitations in existing methods for sentiment analysis of tourism reviews, the authors propose an image-text multimodal sentiment analysis method (TBGAV). It consists of three modules: image sentiment extraction, text sentiment extraction, and image-text fusion. The image sentiment extraction module employs a pre-trained VGG19 model to capture sentiment features. The text sentiment extraction module utilizes the tiny bidirectional encoder representations from transformers (TinyBERT) model, incorporating the bidirectional recurrent neural network and attention (BiGRU-Attention) module for deeper sentiment semantics. The image-text fusion module employs the dual linear fusion approach to correlate image-text links and the maximum decision-making approach for high-precision sentiment prediction. TBGAV achieves superior performance on the Yelp dataset with accuracy, recall rates, and F1 scores of 77.51%, 78.01%, and 78.34%, respectively, outperforming existing methods. Accordingly, TBGAV is expected to help improve travel-related recommender systems and marketing strategies.

KEYWORDS

Attention, BiGRU, Multimodal, Sentiment Analysis, TinyBERT, VGG19

INTRODUCTION

With the rise of social media and online reviews, an increasing number of people are sharing their travel experiences via online travel evaluations (Ahmed et al., 2022; Fotis et al., 2011). These travel reviews contain rich emotional information, and it is crucial for tourism practitioners, travel enthusiasts, and policy makers to accurately understand tourists' emotional experiences and feedback. Therefore, the study of sentiment analysis on travel reviews is of great value and practical significance (Fazzolari & Petrocchi, 2018).

Sentiment analysis is an essential research direction in natural language processing, aiming at recognizing and interpreting sentiment information from text (Chen et al., 2023; Yang et al., 2020).

DOI: 10.4018/IJITWE.334595 *Corresponding Author

By applying sentiment analysis techniques, it is possible to automate the extraction and analysis of a vast quantity of tourism review data to understand tourists' attitudes and emotional tendencies towards different destinations, attractions or services. Such analysis can not only help tourism practitioners gain a thorough understanding of tourists' satisfaction with tourism products and services, but also provide valuable references for destination promotion and improvement.

In recent years, researchers have started applying sentiment analysis to the investigation of tourism evaluations. However, most of the established studies have focused on the analysis of textual data, with little consideration of the image information of the attractions or services described in that review (Nandwani & Verma, 2021). For instance, Li et al. (2020) suggested a deep learning-based sentiment analysis model called dictionary integrated dual-channel convolutional neural network (CNN)-long-short tem memory (LSTM) family of models, which combines the CNN and the LSTM/BiLSTM branches in a parallel fashion. Experiments on multiple difficult datasets (e.g., Stanford Sentiment Treebank) show that the suggested method surpasses a number of baseline approaches; Fang et al. (2022) suggested the utilization of an ELECTRA (i.e., efficient learning of accurate classification of token replacements for encoders) classifier to do sentiment analysis on reviews of tourist attractions. Numerous trials have substantiated the efficacy of the suggested methodology for doing sentiment analysis on reviews of tourist attractions. However, image information in tourist reviews usually contains emotionally relevant visual elements, such as scenery and service quality, which play a significant role in the expression of tourists' emotional experience (Ortis et al., 2020). Therefore, graphical sentiment analysis that integrates image and textual information has higher benefits and accuracy and can provide comprehensive travel review sentiment analysis results (Xu et al., 2021).

Nevertheless, existing sentiment analysis methods based on image and text fusion still face some challenging issues, such as the problem of polysemy in text modality, insufficient contextual information extraction in text modality, difficulty in distinguishing the contribution of comment information to emotional features, and insufficient interaction between image modality and text modality. In order to better solve the multimodal sentiment analysis problem based on tourism review information, by deeply mining the sentiment features in travel reviews, the authors constructed a sentiment classification model integrating image and will evaluate text analysis and its performance through experiments.

The following are the primary contributions of this paper:

1. Tiny bidirectional encoder representations from transformers (TinyBERT) transforms the input text to dynamic word embeddings, which includes positional embeddings, token embeddings, and segment embeddings, thereby resolving the issue of word polysemy while simultaneously lowering the training parameters and the output size.
2. Aiming to maximize the text context relationship, the authors adopt a bidirectional recurrent neural network (BiGRU) structure, and determine the output jointly by the state of the forward hidden layer and the state of the backward hidden layer, resulting in a more accurate classification effect.
3. To effectively differentiate the contribution of user comments to sentiment analysis, the attention mechanism highlights key features by assigning weights to the importance of various comment features.
4. Most of the studies only simply spliced image features and text features into the classifier without fully considering the bidirectional correlation between graphic and text data. Utilizing a dual linear fusion technique, the authors explore the interaction information between image and sentiment data thoroughly.

On Controlling Elasticity of Cloud Applications in CELAR