

Article

An Effective ELECTRA-Based Pipeline for Sentiment Analysis of Tourist Attraction Reviews

Hui Fang ^{1,*}, Ge Xu ¹, Yunfei Long ² and Weimian Tang ¹

¹ College of Computer and Control Engineering, Fujian Provincial Key Laboratory of Information Processing and Intelligent Control, Minjiang University, Fuzhou 350108, China

² School of Computer Science and Electronic Engineering, University of Essex, Colchester CO2 8JT, UK

* Correspondence: fh_mju@126.com; Tel.: +86-0591-83761607

Abstract: In the era of information explosion, it is difficult for people to decide on a tourist destination quickly. Online travel review texts provide valuable references and suggestions to assist in decision making. However, tourist attraction reviews are primarily informal and noisy. Most works in this field focus on shallow machine learning models or non-pretrained deep learning models. These approaches struggle to generate satisfactory classification results. To solve this issue, the paper proposes a pipeline model. In the first step of this paper, we preprocess tourist attraction reviews by performing stopword removal, special character removal, redundancy deletion and negation substitution to reduce noise. Then, we propose an ELECTRA (Efficiently Learning an Encoder that Classifies Token Replacements Accurately) classifier for sentiment analysis of tourist attraction review. Finally, we compare our pipeline model with several representative deep text classification models. Extensive experiments have demonstrated the effectiveness of our approach to sentiment analysis of tourist attraction reviews. We not only provide one high-quality dataset for tourist attraction reviews, but our work can also expand and promote the development of sentiment analysis in other domains.

Keywords: ELECTRA-based; sentiment analysis; text pre-processing; tourist attraction reviews



Citation: Fang, H.; Xu, G.; Long, Y.; Tang, W. An Effective ELECTRA-Based Pipeline for Sentiment Analysis of Tourist Attraction Reviews. *Appl. Sci.* **2022**, *12*, 10881. <https://doi.org/10.3390/app122110881>

Academic Editor: José António Correia

Received: 23 September 2022

Accepted: 24 October 2022

Published: 27 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent years, review texts have provided a novel source of data for travel research. Review texts contain insightful feedback that is spontaneously provided by users. Online travel review texts are naturally considered as one of the most essential sources of user opinion. It is widely understood that reputation plays an important role for informing and affecting user decisions. User-generated Internet content, such as reviews, facilitates this process. Therefore, it is very important to analyze user opinions from data such as online reviews. The growing number of user reviews is the most readily available source of the opinions of crowds. However, the analysis of such opinions from reviews, especially from websites with a large number of users worldwide, is challenging.

Sentiment analysis has been introduced to discover knowledge through user reviews. Sentiment analysis in review texts is a rapidly growing field of study and application, and has been applied in the domain of tourism as well. However, compared to many other domains of sentiment analysis, online travel reviews are usually short texts, written informally—meaning the writer uses slang, misspelled words, and emoticons. Travel is a low-frequency activity, and travel frequency among users approximates Zipf's law from a statistical perspective. The sparsity of noisy short texts, coupled with uneven sentiment distribution, makes it difficult to obtain ideal sentiment analysis results. Although several works have recognized these problems and treated the sentiment analysis of travel review texts as a domain specific classification problem [1–4], they use either shallow machine learning models that lack the ability to extract deep semantic features or use non-pretrained deep learning models that heavily depend on a large amount of labeled text. It is still possible to improve the classification results of sentiment analysis of travel review texts.

Our research objectives mainly focus on the following two aspects:

First, we plan to carry out a series of pre-processing procedures to transform tourist attraction reviews into text that is more appropriate for learning emotional characteristics to ensure the performance gain of learning models. Second, we aim to use a pre-trained model, namely, ELECTRA, as a smart classifier to work on a sentiment analysis task based on tourist attraction reviews together with pre-processing procedures. This is due to ELECTRA's more efficient pre-training process and better performance on small data sets. To verify our approach, we build an expert-annotated review dataset in Chinese related to tourist attractions.

Overall, we provide resources and benchmark model for those who use NLP technology to create sentiment analysis of tourist attraction review.

In summary, this paper has the following contributions:

- We propose a pre-trained pipeline model for the sentiment analysis of tourist attraction reviews, which can exploit the gains of pre-trained language models and which outperforms other baseline models.
- We develop annotation specifications and manually construct a Chinese tourist attraction review dataset to fill the research gap.
- We conduct a detailed comparison between our model and other baseline models in terms of performance evaluation and model evaluation, with an additional ablation experiment. The discoveries from these analyses can promote the development of sentiment analysis of other reviews.

2. Background and Related Works

2.1. Travel Review Sentiment Analysis

Text sentiment analysis (TSA) is the process of extracting users' opinions, sentiments and demands from unstructured subjective texts in a specific domain and distinguishing their polarity. Existing TSAs roughly fall into three main categories: sentiment lexicon and rule-based methods [5–7]; traditional machine learning-based methods [8–11]; and deep learning-based methods [12–20]. Category I has proved to perform poorly when the texts are rich in new words, contextual words or multilingual words. Category II focuses on extracting sentiment features and the combination of different classifiers (e.g., KNN, SVM and Naïve Bayes, etc.). Without fully utilizing the contextual information of the text, their classification accuracy is affected to a certain extent. In order to obtain better classification results, category III introduces deep neural networks—such as CNN, RNN and LSTM networks, Attention networks and Transformer networks—to automatically extract sentiment features and make good use of contextual semantic information.

In recent years, research on sentiment analysis of review texts in tourism has started to receive more and more attention. It is driven by helping tourism stakeholders to better understand and to more quickly grasp relevant tourism information, providing a reliable basis for their decision-making process. Sentiment analysis is considered a key step in restaurant or hotel recommendations for tourists in many works [21–23]. However, only a handful of studies on the sentiment analysis of attraction review can be found [24–26]. It is worth noting that there is no existing work applying a pre-trained model for sentiment analysis of tourist attraction review texts. Pre-trained language models (PLMs) can achieve comparable or even SOTA (State-of-the-Art) results on NLP tasks with a small number of supervised corpora.

2.2. Pre-Trained Language Models

Recently, PLMs, which consist of an extensive neural network previously trained on a large amount of unlabeled data and fine-tuned on downstream tasks, have achieved outstanding performance in several natural language understanding tasks. In the text classification task, Howard and Ruder [27] put forward Universal Language Model Fine-Tuning (ULMFiT) and achieved SOTA results. The encoder part of an encoder–decoder architecture based on deep transformer, such as Generative Pre-trained Transformer (GPT) [28] and

BERT [29], is nowadays one of the most popular task-specific models. In particular, many of the works [15,16,20] proposed using BERT and its variants for sentiment analysis with excellent results. However, BERT and its variants have some fatal flaws: slow convergence of model training; high computational effort; and some inconsistency in inputs during pre-training and fine-tuning. To address these, ELECTRA [30], which uses a different pre-training method acting as a discriminator rather than a generator, has been used. In some tasks, such as similarity comparisons [31] and sequence annotation [32], ELECTRA has been shown to have better performance than BERT.

2.3. Text Pre-Processing

Text pre-processing, especially for informal texts, is an integral step in sentiment analysis and PLMs. The pre-processing that may be involved includes tokenization, part-of-speech tagging, stemming, lemmatization, text cleaning [33], text clarity [34], tagging [35], lexical-grammatical check, spellchecking, stopword removal and negation handling [36]. Pre-processing has a direct impact on the performance of sentiment classification. For instance, Reference [33] indicates that the inappropriate processing of negations leads to biases and misclassification of sentiments. Reference [34] proposes cleaning and normalizing data, negation handling and intensification handling to improve sentiment classification performance. Several studies have shown that pre-processing also contributes to the performance of pre-trained models. Reference [35] leverages lexical simplification to effectively improve the performance of PLMs in text classification. Pre-processing has been verified to solve the limitations of word embedding for affective tasks [36]. Nowadays, the pre-processing of sentiment analysis mainly focuses on English text, and there are only a few works for Chinese text. Reference [37] proposes a method based on Chinese characters rather than words to address the problem of requiring complex pre-processing steps in Chinese text sentiment analysis. Reference [38] studies text pre-processing in Chinese, such as document segmentation, word segmentation and text representation, but only to unify the format of documents before text classification.

3. Methodology

In this section, we introduce our proposed approach to classify tourist attraction review and perform sentiment analysis, given the large number of tourist attraction reviews with significantly different sentiment tendencies and the presence of many meaningless or inauthentic contents, as well as the extensive use of informal terms. These characteristics mean the task of tourist attraction classification requires necessary pre-processing and efficient classifiers. It is essentially a two-step ELECTRA-based pipeline, as shown in Figure 1. In detail, the first step of the pipeline applies a series of pre-processing procedures to convert travel reviews from Ctrip (<https://www.ctrip.com/>, accessed on 15 July 2022) into filtered text, while the second step places the data processed into the classification system based on an ELECTRA language model that has been pre-trained on plain text corpora. In particular, our proposed pre-processing procedures are outlined and the architecture of the classification system adopted is explained.

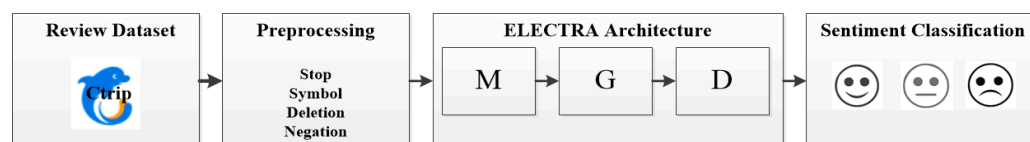


Figure 1. Pipeline overview. M is short for mask operation, and G and D are the generator and discriminator, respectively.

3.1. Pre-Processing Procedures

Collecting raw travel reviews from the Ctrip Travel Website (the largest Online Travel Agency in China) using a Web crawler generally results in a very noisy dataset due to the spontaneity and creativity of the posted comments. Since tourist attraction reviews reflect

the satisfaction of tourists and their evaluations of the quality of services, perceptions of the image of attractions or information provision, they are filled with many modal words and emotional words, coupled with other noisy sources, such as phone numbers, amounts of money, times, dates, addresses and questions.

Our pre-processing procedures include the following four sub-procedures:

1. Stopword Removal

Stopwords (abbreviated as Stop) are the most common words typically filtered out before the classification task. Thus, we removed all the stopwords. Here, we use the “remove-stopwords” method in the Gensim (<https://radimrehurek.com/gensim/index.html>), accessed on 24 July 2022) Library with a stopwords dictionary that integrates Chinese words from the HIT (Harbin Institute of Technology), Baidu and SCU (Sichuan University) thesauruses. Given that the removal of stopwords has a significant impact on the set of feature vectors required for classification and the classification effect [39], an optimal selection of stopwords is required. It is not preferable to have a stopwords dictionary with as many stopwords as possible, but rather, it is better to have a targeted dictionary. Since the combination of words such as exclamations, onomatopoeia and pronouns with other words in emotional texts often has a strong emotional tendency, all words other than punctuation and non-semantic words are retained in the stopwords dictionary as much as possible. The final stopwords dictionary contains 1336 stopwords. We will make the final stopwords dictionary publicly accessible for the research community.

2. Special Character Removal

Special characters (abbreviated as Symbol) such as . , ()[]{} should be removed in order to eliminate differences when assigning polarity. Here, we use the “remove_stopwords” method in the Gensim Library with a symbol dictionary that integrates special characters from the HIT thesaurus. Gensim is an open source software library that uses modern statistical machine learning and is designed to process large collections of text using data streams and incremental online algorithms, unlike most other machine learning packages that target only in-memory processing. The final symbol dictionary contains 263 special characters.

3. Redundancy Deletion

Repeated statements may disrupt polarity distribution, and phone numbers, amounts of money, times, dates and questions do not contribute to sentiment tendencies. Thus, we delete these statements and phrases, abbreviated as Deletion. We identify redundant statements from the text using the NLTK `regexprtokenizer` and delete them. Several regular matching examples are illustrated in Table A1 in the Appendix.

4. Negation Substitution

Negation substitution (abbreviated as Negation) plays a critical role, as negation words would invert the word or sentence polarity in sentiment analysis. Thus, we substitute the negation and the negated word with its antonym. First, we identify the negation words in tokenized text using a negation dictionary (contains 243 negation words). Then, the antonym of the token following the negated word is looked up in the antonym dictionary (contains 18,797 pairs). If an antonym is found, the negation word and the negated word are replaced with the antonym. For example, we replace 不开心 (not happy) with 沮丧 (depressed).

3.2. ELECTRA System Architecture

While BERT and its variants produce excellent results on downstream NLP tasks, they require a large amount of computation to be effective. This is because such masked language models (MLM) mask a few words randomly during the training process and predict very limited words. Furthermore, because of the arbitrary choice of the masked token, it would be challenging to learn as many meaningful tokens as possible, such as emotional words and opinion words, for the sentiment analysis task. As an alternative,

ELECTRA proposes a more efficient pre-training task that can compensate well for these shortcomings, so we introduce it as a sentiment classification model for attraction review to achieve a more satisfying classification performance. The ELECTRA system architecture is shown in Figure 2.

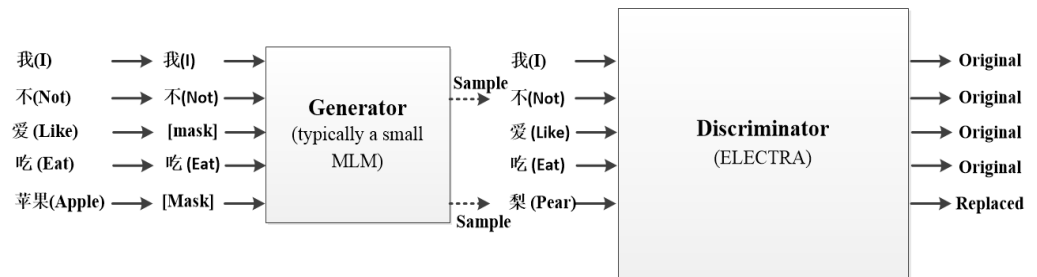


Figure 2. ELECTRA architecture overview.

To go into more detail, just like Generative Adversarial Networks (GAN), the ELECTRA architecture consists of two networks: a generator and a discriminator. Both parts use transformer-based encoding networks to obtain the vector representation of the input word sequence. In the Chinese dataset, as Chinese characters are different from English, which is formed by alphabet-like symbols, the word tokenizer uses the traditional Chinese Word Segmentation (CWS) tool to split the text into several words instead of small fragments. In this way, whole word masking in Chinese could be adopted to mask the word instead of individual Chinese characters. In the generator, the goal is to train a masked language model. Its structure is similar to BERT, i.e., given an initial input sequence = {我 (I), 不 (not), 爱 (like), 吃 (eat), 苹果 (apple)}, the Chinese words 爱 (like) and 苹果 (apple) in the masked sequence are first replaced by [MASK] according to a certain percentage to obtain the generator’s input. The process can be formulated as:

$$x^{masked} = REPLACE(x, index, [MASK]), \tag{1}$$

where $index = [id_1, \dots, id_k]$ is the index sequence for selected positions.

Then, a vector representation is obtained through a generative network, typically a small MLM. Followed by a softmax layer, a sample word \hat{x}_i is predicted for the location of the mask in the generator’s input:

$$\hat{x}_i \sim p_G(x_i | x^{masked}) \tag{2}$$

The objective function of training is to maximize the likelihood of the masked words. The prediction result replaces the original masked word, e.g., the Chinese words 爱 (like) and 苹果 (apple) are replaced by 爱 (like) and 梨 (pear), respectively:

$$x^{corrupt} = REPLACE(x, index, \hat{x}), \tag{3}$$

where $index$ is the same as above.

In the discriminator, a new pre-training task known as “Replaced Token Detection (RTD)” is applied. More specifically, a discriminative model is trained to predict whether the word at each position of the discriminator’s input sequence has been replaced by the generator, with “Original” or “Replaced” as the classification result. Here, only the word 苹果 (apple) can be found to have changed.

The generator’s loss function $L_{MLM}(x, \theta_G)$ can be formulated as:

$$x^{corrupt} = REPLACE(x, index, \hat{x}), \tag{4}$$

and the discriminator's loss function $L_{Disc}(x, \theta_D)$ can be formulated as:

$$L_{Disc}(x, \theta_D) = E \left(\sum_{t=1}^n -1(x_t^{corrupt} = x_t) \log D(x^{corrupt}, t) - 1(x_t^{corrupt} \neq x_t) \log(1 - D(x^{corrupt}, t)) \right) \quad (5)$$

The final loss function is the weighted sum of two loss functions:

$$\min_{\theta_G, \theta_D} \sum_{x \in X} L_{MLM}(x, \theta_G) + \lambda L_{Disc}(x, \theta_D), \quad (6)$$

where λ is a weighting factor and X is a large raw text corpus. After pre-training, we use the discriminator to fine-tune the sentiment analysis task.

The process of replacement identification using a discriminator converts a prediction problem into a binary classification one that is characterized by allowing words in all positions to be predicted, resulting in an increase in efficiency and faster convergence. This new pre-training task is more effective than MLM because the model learns from all input words, not just from a small subset of masked ones. The contextual representation learned by ELECTRA is substantially better than that learned by MLM such as BERT in the same model size, data and computational conditions, with particularly large gains from small datasets.

4. Experimental Design

4.1. Dataset

There is a large number of datasets publicly available to facilitate the study of sentiment analysis in recent years. Most of these datasets are movie or product reviews, such as from Yelp (<https://www.kaggle.com/yelp-dataset/yelp-dataset>, accessed on 16 July 2022), Amazon (<https://www.kaggle.com/datafiniti/consumer-reviews-of-amazon-products>, accessed on 16 July 2022), IMDb (<https://www.kaggle.com/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews>, accessed on 16 July 2022) and SST (<https://www.kaggle.com/atulanandjha/stanford-sentiment-treebank-v2-sst2>, accessed on 16 July 2022). The tourism-related dataset only covers restaurants or hotels, mostly in English, such as SemEval2014-ABSA-Restaurant-Reviews (<https://alt.qcri.org/semeval2014/task4>, accessed on 16 July 2022) and ChusentiCorp-h1 (<https://raw.githubusercontent.com/SophonPlus/ChineseNLPCorpus>, accessed on 16 July 2022). Therefore, we have constructed a sentiment dataset of tourist attraction reviews in Chinese from Ctrip. Our dataset is called SenTARev (Sentiment analysis for Tourist Attraction Reviews). In the adopted labeling scheme, there are two types of labels, positive (PL) and negative (NL). There are three types of affective outcomes, negative (Neg), positive (Pos) and neutral (Neu). Considering labeling as a classification task, it is defined as follows: if positive sentiment is detected, PL outputs a value of 1 as a positive class and 0 otherwise; if negative sentiment is detected, NL outputs a value of 1 as a negative class and 0 otherwise. Thus, for PL, a value of 1 means a positive or neutral sentiment and a value of 0 means a neutral or negative sentiment. For NL, the value 1 implies a negative or neutral sentiment and the value 0 implies a neutral or positive sentiment. We organize annotators to perform polarity annotation and data inspection on the raw data, which is finally used as a supervised corpus for fine-tuning the ELECTRA model. In the quality-control process of human-annotated data, each attraction was assigned an annotator and an inspector. We set a unified annotation specification before annotation to ensure the consistency of the data. Every annotator had to perform real-time inspection, and every inspector had to complete full sample inspection and sampling inspection. Table 1 illustrates our annotation scheme and SenTARev's label distribution.

Table 1. SenTARev’s Label Distribution.

Combination		Sentiment Result	Train	Validation	Test
PL	NL				
0	0	Neu	638	212	213
1	1	Neu			
0	1	Neg	1164	388	388
1	0	Pos	16,618	5540	5539

4.2. Baseline

To evaluate the effectiveness of the proposed approach, we compared it with several baselines. Each baseline is a pipeline model that includes the same pre-processing procedures mentioned above and the following classifier:

- TextCNN [40]: a classical CNN text classifier;
- TextRNN [41]: a classical RNN text classifier. It adopts BiLSTM to learn text representation;
- TextRCNN [42]: an RNN text classifier. It adopts BiLSTM with a pooling mechanism to learn text representation;
- TextRNN-Att [43]: an RNN-based text classifier. It adopts BiLSTM with an attention mechanism to learn text representation;
- BERT [29]: the representative of Masked Language Modeling (MLM) pre-trained models. A linear classification layer on top of BERT’s output;
- RoBERTa [44]: a rigorously optimized BERT model. A linear classification layer on top of RoBERTa’s output.

4.3. Evaluation Metrics

In this paper, we adopt precision (P), recall (R) and f1-score (F) as the main evaluation metrics for sentiment analysis performance. They can be respectively calculated as:

$$P_p = \text{correct}_p / \text{assigned}_p, \quad (7)$$

$$R_p = \text{correct}_p / \text{total}_p, \quad (8)$$

$$F_p = 2P_p R_p / (P_p + R_p), \quad (9)$$

where p where p denotes the polarity, i.e., Neu, Neg and Pos. We also utilize macro-average (m_{avg}) and weighted average (w_{avg}) as comprehensive evaluation metrics of the deep learning model. These calculations are shown as:

$$m_{avg_X} = \sum_p X_p / 3, \quad (10)$$

$$w_{avg_X} = \sum_p X_p * \text{sup}_p / \sum_p \text{sup}_p, \quad (11)$$

where X denotes P , R or F , and sup denotes the number of support samples.

4.4. Model Training

Under the Hugging Face framework, the ELECTRA model used is made available by the Joint Laboratory of the HIT and iFLYTEK Research team. The Chinese ELECTRA is pre-trained on two corpora: the first corpus is the source data consisting of the Chinese Wikipedia dump, while the second corpus is further extended with data from encyclopedia, news and question and answering websites, which has 5.4 billion words and is over ten times bigger than the Chinese Wikipedia. Fine-tuning of the model was performed by using labelled reviews in the training set of the SenTARev dataset. Categorical cross entropy was used as the loss function during training, and the fully connected classification layer was learned correspondingly. For the fine-tuned ELECTRA, the hyper-parameters used are shown in Table A2 in the Appendix A.

4.5. Result and Discussion

The comparison results of performance evaluation on the SenTraRev dataset are presented in Table 2. For each method, the results are obtained on the best model. As shown in Table 2, firstly, we can see that the classification performance of our ELECTRA-based pipeline model is better than all selected classical text classification models in precision, recall and f1-score by a significant margin. Then, we observe that, with complex pre-trained targets and large model parameters, large-scale PLMs can effectively capture knowledge from large amounts of labeled and unlabeled data. By storing knowledge using a large number of parameters and fine-tuning it for specific tasks, the rich knowledge implicitly encoded by a large number of parameters can benefit the sentiment analysis task. Hence, from Table 2, we can observe that the pre-trained models have outperformed other models. Thirdly, our ELECTRA-based pipeline model also achieves better performance compared with other pre-trained models on all measurements. This is due to its more efficient pre-training task. This novel pre-training task makes our ELECTRA-based text classification model learn deeper contextual sentiment features in reviews than other pre-trained based models. In terms of text classification, arbitrary masking patterns make it difficult for BERT-based models to learn all meaningful information about sentiment tendencies. This is especially true in the RoBERTa-based model, whose dynamic masking further undermines the effectiveness of learning sentiment tokens, making performance even worse than CNN/RNN models. These have further blurred the boundary between biased polarity and unbiased polarity during training and increased the difficulty of classification. The subtle masking design makes our ELECTRA-based model obtain better generalization performance on text classification.

Table 2. Experimental results of performance comparison.

Models	F_{Neu}	F_{Neg}	F_{Pos}	m_{avgF}	w_{avgF}
TextCNN	0.4074	0.7945	0.9759	0.7259	0.9443
TextRNN	0.4010	0.8229	0.9798	0.7345	0.9494
TextRCNN	0.4802	0.8571	0.9832	0.7735	0.9574
TextRNN-Att	0.4504	0.8722	0.9822	0.7683	0.9565
BERT	0.5410	0.8903	0.9851	0.7756	0.9594
RoBERTa	0.5155	0.8781	0.9761	0.7894	0.9418
ELECTRA	0.5818 *	0.9062 *	0.9864 *	0.8246 *	0.9663 *

* Denotes significance at $p \leq 0.05$.

4.6. Ablation Studies

For further evaluation, an ablation study was performed. This involved firstly discarding the pre-processing procedure, then only enabling one pre-processing sub-procedure, then enabling any two sub-procedures and disabling one pre-processing sub-procedure, and finally retaining other sub-procedures. The results of ablation studies are detailed in Table 3. A primary goal of this work is to identify the most effective sub-procedure for PLMs for sentiment analysis. Observing the results of the individual sub-procedure on the SenTARev dataset, it is worth noting that even a single pre-processing sub-procedure can bring improvements. Among these four pre-processing sub-procedures, negation substitution appears to be the most effective, verifying its importance in sentiment classification. Then we find redundancy deletion and special character removal also contribute to improvement. Stopwords removal had minimal impact. Performance with negation pre-processing is generally better than without it when the same number of pre-processing sub-procedures is included. Although it does not present a multiplicative or proportional relationship, performance becomes better as the number of pre-processing sub-procedure increases. We note that the best performance comes from combining all the pre-processing procedures.

Table 3. Ablation experimental results on performance comparison.

Models	F_{Neu}	F_{Neg}	F_{Pos}	m_avg_F	w_avg_F
None	0.4954	0.8883	0.8029	0.8022	0.9425
Stop	0.5000	0.8856	0.9145	0.8041	0.9463
Symbol	0.5158	0.8857	0.9269	0.8065	0.9475
Deletion	0.5118	0.8882	0.9274	0.8058	0.9488
Negation	0.5294	0.8833	0.9387	0.8108	0.9501
Stop + Symbol	0.5326	0.8850	0.9464	0.8110	0.9521
Deletion + Symbol	0.5405	0.8876	0.9498	0.8120	0.9513
Deletion + Stop	0.5422	0.8875	0.9421	0.8102	0.9502
Symbol + Negation	0.5569	0.8964	0.9486	0.8140	0.9543
Deletion + Negation	0.5589	0.8977	0.9462	0.8143	0.9573
Stop + Negation	0.5400	0.8947	0.9438	0.8122	0.9548
Deletion + Symbol + Negation	0.5799	0.8982	0.9848	0.8232	0.9646
Deletion + Symbol + Stop	0.5521	0.8967	0.9768	0.8215	0.9619
Deletion + Stop + Negation	0.5755	0.8943	0.9744	0.8207	0.9622
Stop + Symbol + Negation	0.5736	0.8941	0.9713	0.8218	0.9638
ALL	0.5818 *	0.9062 *	0.9864 *	0.8246 *	0.9663 *

* Denotes significance at $p \leq 0.05$.

5. Conclusions

In this paper, we set out to improve sentiment analysis in tourist review data. Firstly, we constructed a Chinese tourist attraction review dataset, SenTARev, from Ctrip. Then, we proposed a two-step pipeline approach for the sentiment analysis of tourist attraction reviews. We found that an ELECTRA-based pipeline model is highly efficient at the sentiment analysis of tourist attraction reviews in terms of model performance. This represented, on average, a 6.33 percent improvement in m_avg_F and averaged a 1.5 percent improvement in w_avg_F . In addition, we also found that pre-processing can further enhance pre-trained sentiment classification models, especially negation substitution. Our work can expand and promote the development of sentiment analysis of other domains. In the future, we will explore pre-trained models with travel knowledge enhancement to improve their ability to understand and represent domain knowledge. Furthermore, we will integrate pre-processing with deep learning models through transformation and filtering with the chi-quared method to further improve the efficiency of Chinese sentiment analysis.

Author Contributions: Conceptualization, H.F. and G.X.; methodology, H.F.; software, W.T.; validation, H.F., Y.L. and W.T.; formal analysis, H.F.; investigation, H.F.; resources, G.X.; data curation, W.T.; writing—original draft preparation, H.F.; writing—review and editing, Y.L.; visualization, W.T.; supervision, G.X.; project administration, G.X.; funding acquisition, G.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Central Leading Local Project “Fujian Mental Health Human-Computer Interaction Technology Research Center” (grant number 2020L3024).

Institutional Review Board Statement: Not acceptable.

Informed Consent Statement: Not acceptable.

Data Availability Statement: The GitHub address for the SenTARev dataset is <https://github.com/peterpan23/SenTARev->, accessed on 11 August 2022.

Acknowledgments: The authors express our acknowledgement to my colleagues in Minjiang University and Assistant Professor YUNFEI LONG in University of Essex for their valuable suggestions.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

13. Luo, F.; Wang, H. Chinese text sentiment classification by H-RNN-CNN. *Beijing Da Xue Xue Bao* **2018**, *54*, 459–465.
14. Yang, C.; Zhang, H.; Jiang, B.; Li, K. Aspect-based sentiment analysis with alternating coattention networks. *Inf. Process. Manag.* **2019**, *56*, 463–478. [[CrossRef](#)]
15. Araci, D.F.; Genc, Z. Financial sentiment analysis with pre-trained language models. *arXiv* **2019**, arXiv:1908.10063.
16. Azzouza, N.; Akli-Astouati, K.; Ibrahim, R. Twitterbert: Framework for twitter sentiment analysis based on pre-trained language model representations. In Proceedings of the International Conference of Reliable Information and Communication Technology, Johor, Malaysia, 22–23 September 2019; Springer: Cham, Switzerland; pp. 428–437.
17. Jelodar, H.; Wang, Y.; Orji, R.; Huang, S. Deep sentiment classification and topic discovery on novel coronavirus or COVID-19 online discussions: NLP using LSTM recurrent neural network approach. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 2733–2742. [[CrossRef](#)]
18. Wei, J.; Liao, J.; Yang, Z.; Wang, S.; Zhao, Q. BiLSTM with multi-polarity orthogonal attention for implicit sentiment analysis. *Neurocomputing* **2020**, *383*, 165–173. [[CrossRef](#)]
19. Mohamed, E.H.; Moussa, M.E.; Haggag, M.H. An enhanced sentiment analysis framework based on pre-trained word embedding. *Int. J. Comput. Intell. Appl.* **2020**, *19*, 2050031. [[CrossRef](#)]
20. Pota, M.; Ventura, M.; Catelli, R.; Esposito, M. An effective BERT-based pipeline for Twitter sentiment analysis: A case study in Italian. *Sensors* **2020**, *21*, 133. [[CrossRef](#)] [[PubMed](#)]
21. Petrusel, M.R.; Limboi, S.G. A restaurants recommendation system: Improving rating predictions using sentiment analysis. In Proceedings of the 2019 21st International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), Timisoara, Romania, 4–7 September 2019; pp. 190–197.
22. Ray, B.; Garain, A.; Sarkar, R. An ensemble-based hotel recommender system using sentiment analysis and aspect categorization of hotel reviews. *Appl. Soft Comput.* **2021**, *98*, 106935. [[CrossRef](#)]
23. Asani, E.; Vahdat-Nejad, H.; Sadri, J. Restaurant recommender system based on sentiment analysis. *Mach. Learn. Appl.* **2021**, *6*, 100114. [[CrossRef](#)]
24. An, H.W.; Moon, N. Design of recommendation system for tourist spot using sentiment analysis based on CNN-LSTM. *J. Ambient Intell. Humaniz. Comput.* **2019**, *13*, 1653–1663. [[CrossRef](#)]
25. Luo, Y.; Zhang, X.; Qin, Y.; Yang, Z.; Liang, Y. Tourism attraction selection with sentiment analysis of online reviews based on probabilistic linguistic term sets and the IDOCRIW-COCOSO model. *Int. J. Fuzzy Syst.* **2021**, *23*, 295–308. [[CrossRef](#)]
26. Qin, Y.; Wang, X.; Xu, Z. Ranking tourist attractions through online reviews: A novel method with intuitionistic and hesitant fuzzy information based on sentiment analysis. *Int. J. Fuzzy Syst.* **2021**, *24*, 755–777. [[CrossRef](#)]
27. Howard, J.; Ruder, S. Universal language model fine-tuning for text classification. *arXiv* **2018**, arXiv:1801.06146.
28. Radford, A.; Narasimhan, K.; Salimans, T.; Sutskever, I. Improving Language Understanding by Generative Pre-Training. 2018. Available online: <https://www.semanticscholar.org/paper/Improving-Language-Understanding-by-Generative-Radford-Narasimhan/cd18800a0fe0b668a1cc19f2ec95b5003d0a5035>, (accessed on 2 July 2022).
29. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
30. Clark, K.; Luong, M.T.; Le, Q.V.; Manning, C.D. Electra: Pre-training text encoders as discriminators rather than generators. *arXiv* **2020**, arXiv:2003.10555.
31. Ostendorff, M.; Ruas, T.; Blume, T.; Gipp, B.; Rehm, G. Aspect-based document similarity for research papers. *arXiv* **2020**, arXiv:2010.06395.
32. Ding, J.; Liu, X. Named Entity Recognition Model of Telecommunication Network Fraud Crime Based on ELECTRA-CRF. *Inf. Netw. Secur.* **2020**, *6*, 63–69.
33. Mukherjee, P.; Badr, Y.; Doppalapudi, S.; Srinivasan, S.M.; Sangwan, R.S.; Sharma, R. Effect of negation in sentences on sentiment analysis and polarity detection. *Procedia Comput. Sci.* **2021**, *185*, 370–379. [[CrossRef](#)]
34. Duong, H.T.; Nguyen-Thi, T.A. A review: Preprocessing techniques and data augmentation for sentiment analysis. *Comput. Soc. Netw.* **2021**, *8*, 1–16. [[CrossRef](#)]
35. Tang, D.; Wei, F.; Yang, N.; Zhou, M.; Liu, T.; Qin, B. Learning sentiment-specific word embedding for twitter sentiment classification. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, Baltimore, AR, USA, 22–27 June 2014; pp. 1555–1565.
36. Agrawal, A.; An, A.; Papagelis, M. Learning emotion-enriched word representations. In Proceedings of the 27th International Conference on Computational Linguistics, Santa Fe, NM, USA, 20–26 August 2018; pp. 950–961.
37. Lan, Q.; Ma, H.; Li, G. Characters-based sentiment identification method for short and informal Chinese text. *Inf. Discov. Deliv.* **2018**, *46*, 57–66. [[CrossRef](#)]
38. Yu, Y.; Yin, Y. Research on Chinese Text Sentiment Classification Process. In Proceedings of the 3rd International Conference on Mechatronics Engineering and Information Technology (ICMEIT 2019), Hanzhou, China, 15 April 2019; Atlantis Press: Beijing, China; pp. 899–908.
39. Qin, G.; Sanhong, D.; Hao, W. Chinese stopwords for text clustering: A comparative study. *Data Anal. Knowl. Discov.* **2017**, *1*, 72–80.
40. Kim, Y. Convolutional neural networks for sentence classification. In Proceedings of the EMNLP 2014 Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, 25–29 October 2014; pp. 1746–1751.

41. Liu, P.; Qiu, X.; Huang, X. Recurrent neural network for text classification with multi-task learning. *arXiv* **2016**, arXiv:1605.05101.
42. Lai, S.; Xu, L.; Liu, K.; Zhao, J. Recurrent convolutional neural networks for text classification. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, Austin, TX, USA, 25–30 January 2015.
43. Zhou, P.; Shi, W.; Tian, J.; Qi, Z.; Li, B.; Hao, H.; Xu, B. Attention-based bidirectional long short-term memory networks for relation classification. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Berlin, Germany, 7–12 August 2016; pp. 207–212.
44. Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. Roberta: A robustly optimized bert pretraining approach. *arXiv* **2019**, arXiv:1907.11692.