Research on Sentiment Classification of Online Travel Review Text

Wen Chen [1,2], Zhiyun Xu [1], Xiaoyao Zheng [3], Qingying Yu [3] and Yonglong Luo [1,3,*]

[1] School of Geography and Tourism, Anhui Normal University, Wuhu 241003, China; wchen@ahnu.edu.cn (W.C.); zhyxu@mail.ahnu.edu.cn (Z.X.)
[2] School of Mathematics and Computer Science, Tongling College, Tongling 244000, China
[3] Anhui Provincial Key Laboratory of Network and Information Security, Anhui Normal University, Wuhu 241003, China; zxiaoyao@ahnu.edu.cn (X.Z.); ahnuyuq@ahnu.edu.cn (Q.Y.)
* Correspondence: ylluo@ustc.edu.cn; Tel.: +86-0553-5910645

check for updates

Abstract: In recent years, the number of review texts on online travel review sites has increased dramatically, which has provided a novel source of data for travel research. Sentiment analysis is a process that can extract tourists' sentiments regarding travel destinations from online travel review texts. The results of sentiment analysis form an important basis for tourism decision making. Thus far, there has been minimal concern as to how sentiment analysis methods can be effectively applied to improve the effect of sentiment analysis. However, online travel review texts are largely short texts characterized by uneven sentiment distribution, which makes it difficult to obtain accurate sentiment analysis results. Accordingly, in order to improve the sentiment classification accuracy of online travel review texts, this study transformed sentiment analysis into a multi-classification problem based on machine learning methods, and further designed a keyword semantic expansion method based on a knowledge graph. Our proposed method extracts keywords from online travel review texts and obtains the concept list of keywords through Microsoft Knowledge Graph. This list is then added to the review text to facilitate the construction of semantically expanded classification data. Our proposed method increases the number of classification features used for short text by employing the huge corpus of information associated with the knowledge graph. In addition, this article introduces online travel review text preprocessing, keyword extraction, text representation, sampling, establishment classification labeling, and the selection and application of machine learning-based sentiment classification methods in order to build an effective sentiment classification model for online travel review text. Experiments were implemented and evaluated based on the English review texts of four famous attractions in four countries on the TripAdvisor website. Our experimental results demonstrate that the method proposed in this paper can be used to effectively improve the accuracy of the sentiment classification of online travel review texts. Our research attempts to emphasize and improve the methodological relevance and applicability of sentiment analysis for future travel research.

Keywords: user generated content; sentiment analysis; classification; keyword extraction; text representation; sampling; machine learning; TripAdvisor

1. Introduction

Tourism research has entered the era of big data. Based on big data analysis, academia and industry are now better positioned to understand and explore tourist behavior and the tourism market. Li et al. [1] contend that big data analysis can provide sufficient data without introducing sampling bias, and can also make up for the sample size limitations encountered by the survey data, thereby

enabling a better understanding of tourist behavior. Sivarajah et al. [2] argued that big data analysis can lead to new knowledge; subsequently, such analysis has become the mainstream method used to obtain useful information.

From blogs and social media posts to online travel review sites, user-generated content (UGC) is one of the most important data sources for big data. UGC comprises insightful feedback that is spontaneously provided by users. This feedback information is widely available at little to no cost and can also be easily obtained [3]. Such feedback also has potential commercial value in fields such as targeted advertising, customer–company relationships, and brand communication [4,5].

Online travel review sites, such as TripAdvisor, generate large amounts of text-based online travel review data, which constitute an important type of UGC [6]. Online review text data can help researchers and practitioners to correctly understand tourists' travel preferences and needs [7,8]. The opinions expressed in user-generated comments also play an important role in influencing the choices of potential tourists [9,10].

The characteristics of big data have complicated the process of knowledge extraction. The question of how to transform data into valuable knowledge has become crucial for big data applications [11,12]. Previous research into online reviews has mainly focused on the quantitative ratings provided on the website, ignoring the text of online reviews [3]. Ratings cannot provide any information about the specific product characteristics that visitors like or dislike, and such information is typically included in the review text [5,13]. In addition, many users are overwhelmed by the enormous amount of review information provided on travel online review sites. Researchers in other fields have also raised similar questions. Ali et al. [14] noted that while urban traffic congestion is rapidly increasing, a city's rating score is insufficient to provide accurate information; however, comments or tweets may help travelers and traffic managers to understand all aspects of the city. As a result, it is necessary to establish an effective mechanism to help users identify the main content and emotions embedded in the review text [15].

Human emotions and emotional reasoning are understood to be important factors that influence consumer decision-making [16]. This makes sentiment analysis an effective method for mining the connotations of online travel review texts. Text sentiment analysis methods can be divided into dictionary-based methods [17], machine learning methods [13,18], deep learning methods [19,20], and hybrids of the above methods [21,22]. Alaei et al. [8] contend that dictionary-based systems rely on the use of sentiment dictionaries and rule sets. Their article proposes that such methods are unable to adapt to the rapid increase in data volume in the era of big data, so it is necessary to develop more effective automated methods for sentiment analysis. Deep learning methods usually require a large amount of training data to fully realize their potential; this training data usually requires expensive class labeling [23]. Among machine learning methods, support vector machines (SVM) and naive Bayes are the most widely used in the tourism-related sentiment analysis context [13]. Compared with neural networks, SVM and naive Bayes require fewer class annotations to train the model [8]. Most studies on the subject have shown [18] that SVM-based sentiment analysis of text produces superior results relative to other machine learning methods. Kirilenko et al. [13] compared automatic text sentiment analysis classifiers with humans and evaluated whether various types of automatic classifiers are suitable for typical applications in the tourism, hotel, and marketing research contexts. The article argues that on difficult and noisy datasets, automatic classifiers achieve worse performance than humans. It can therefore be concluded that the existing sentiment analysis technology needs to be improved to enable the analysis of specific data.

Contemporary researchers have proposed many effective solutions to improve the performance of SVM in sentiment analysis. Successful feature extraction is one of the main challenges faced by machine learning methods [24]. Feature extraction can reduce information loss and achieve improved discrimination ability in sentiment classification [25] tasks. In their study of feature selection methods, Manek et al. [25] proposed a Gini index feature selection method based on SVM to carry out sentiment classification for a large movie review dataset. Ali et al. [26] proposed a robust classification technology

based on SVM and fuzzy domain ontology (FDO), used for the recognition of comment features and the mining of semantic knowledge. Their experimental results showed that the integration of FDO and SVM greatly improves the accuracy of extracting comments and opinion words, as well as the accuracy of opinion mining. Parlar et al. [27] proposed a new feature selection method based on the query expansion term weighting method in the information retrieval context. This study uses four classifiers to compare their method with other widely used feature selection methods, thereby verifying their method's effectiveness. Zainuddin et al. [28] proposed a latent semantic analysis (LSA) and random projection (RP) feature selection method for the sentiment analysis of Twitter data, and thereby constructed a new Twitter mixed sentiment classification method. Kumar et al. [29] introduced swarm intelligence algorithms into the field of feature optimization in order to improve the sentiment classification performance accuracy. Pu et al. [30] used a variety of features to identify candidate opinion sentences, then used structured SVM to encode these opinion sentences for document sentiment classification. This article resolves the issue of sentiment classification problems arising when the sentiment of most sentences is inconsistent with the sentiment of the document overall.

As an effective feature selection method, semantic expansion has also been widely studied. Adhi et al. [31] designed a sentiment analysis model based on a naive Bayes classifier and the semantic extension method, proving that the semantic extension method can improve the accuracy of sentiment analysis. Fang et al. [32] integrated the context features extracted from the comment sentences and the external knowledge retrieved from the sentiment knowledge graph into a neural network to compensate for the lack of available training data, consequently obtaining better sentiment analysis results. At the same time, as an effective channel for semantic expansion, knowledge bases such as WordNet and ConceptNet are widely used in sentiment analysis in multiple languages. Alowaidi et al. [33] proposed using Arabic WordNet as an external knowledge base to enrich the representation of tweets due to the weakness of the bag of words model; the use of naive Bayes and SVM on the Arab Twitter dataset verified that this external knowledge base can be used to improve sentiment analysis accuracy. Asgarian et al. [34] used Persian WordNet to generate a review corpus, proving that sentiment dictionary quality plays a key role in improving the quality of sentiment classification in the Persian language. Moreover, Agarwal et al. [35] proposed a novel sentiment analysis model based on ConceptNet and common sense extracted from context information.

At the same time, a number of scholars in tourism research have studied the application of sentiment analysis to tourism and hospitality-related data. Several existing works [8,13] have already summarized the sentiment analysis methods adopted by the academic community in the tourism context prior to 2016; therefore, this article only summarizes the relevant literature published after 2017 in Table 1. Among these works, Höpken et al. [36] extracted customer feedback from two online platforms and carried out sentiment analysis and opinion mining, verifying that SVM is best able to solve the problem of sentiment analysis compared with other related methods. Akhtar et al. [37] used topic modeling technology to identify hidden information and other aspects, then performed sentiment analysis on classified hotel review text sentences. Ma et al. [38] performed sentiment analysis on TripAdvisor's review data using Leximancer. Ko et al. [39] applied statistical analysis methods to a large number of consumer review texts obtained from Expedia, enabling these authors to understand the experiences of hotel guests and analyze their association with satisfaction. Stepchenkova et al. [40] selected and compared three of the best-performing sentiment analysis methods to quantify respondents' views on travel in China. Bansal et al. [41] further proposed a sentiment classification method based on mixed attributes. By capturing implicit word relationships and combining domain-specific knowledge, these authors were able to obtain a fine-grained emotional orientation of online consumer reviews. Finally, Lawani et al. [42] used the AFINN dictionary (a lexicon based on unigrams) to extract the sentiments from comments left by Airbnb guests and derive a quality score from those comments.

An analysis of the above literature reveals that the academic community has carried out fruitful work in the field of sentiment analysis, particularly as regards the feature selection of SVM. Although these related topics have been extensively researched, certain specific types of content,

such as online travel review texts for TripAdvisor, still present some challenges when using sentiment analysis [43]. This is because the key features of reviews vary significantly from site to site, meaning that it cannot be assumed that the sentiment analysis method and findings of a certain site will be applicable to all other review sites [44]. On the subject of the sentiment analysis of online travel review texts, most existing sentiment analysis models fail to comprehensively and effectively consider the data characteristics of travel review texts during the modeling process. Online travel review texts have their own inherent characteristics. Most review texts are short, which makes it difficult to extract keywords; in addition, the sentiment distribution of short texts is uneven [45] (for example, the texts with the highest and lowest scores are comparatively few). These characteristics make it difficult for accurate sentiment analysis results to be obtained for online travel review texts [46]. In addition, the accuracy of existing automated sentiment analysis methods is also low [13].

Table 1. The methods of sentiment analysis used in tourism research.

Reference	Methods	Data
[36]	Word list-based methods, supervised learning methods (including k nearest neighbors, support vector machines and naive Bayes)	TripAdvisor, Booking
[37]	Developed by the author	TripAdvisor
[38]	Leximancer	TripAdvisor
[39]	Statistical Analysis	Expedia
[40]	Deeply Moving, Pattern and SentiStrength	Survey data
[41]	Developed by the author	TripAdvisor, Amazon
[42]	AFINN	Airbnb

In order to deal with the sentiment analysis-related challenges brought about by the data features of online travel review texts, this study converted the sentiment analysis of online travel review texts into a multi-classification process based on machine learning methods, and further conducted research on sentiment classification methods for such texts. In order to improve the classification accuracy of online travel review texts, the current research mainly addresses the following problems related to previous research. The main contributions of the paper include:

- Based on the word similarity calculation results, the present study compares three keyword extraction methods and provides the most suitable keyword extraction method for online travel review text.

- After considering the sparse features of online travel review texts, this paper expands the semantics of text keywords based on Microsoft Knowledge Graph in order to build richer and more valuable classification features.

- To address the problem of uneven sentiment distribution in online travel review texts, two types of sampling methods are compared and the most suitable online travel review text sampling method is identified.

- This article introduces the online travel review text preprocessing method, Word2vec-based text representation method, classification label acquisition method, and machine learning method-based sentiment classification method, thereby presenting the entire sentiment analysis process of online travel review texts.

- TripAdvisor is a frequently used text source in sentiment analysis [47]; thus, by analyzing more than 20,000 review text datasets for four famous attractions in four countries derived from TripAdvisor, this paper validates the proposed method from a relatively extensive sample, which allows us to draw more reliable conclusions. Experimental results reveal that the method proposed in this paper is better suited to processing the sentiment classification of online travel review texts, and consequently provides a reference for related travel research.

2. Materials and Methods

Sentiment analysis generally includes multiple steps [48]. As can be seen from Figure 1, the sentiment analysis process proposed in this paper includes the following five steps:
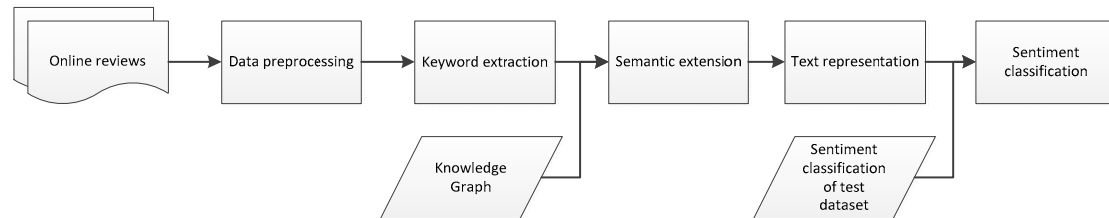

Figure 1. System framework.

(1) Data retrieval. In this study, a crawler program written in Python was used to obtain the texts, namely English descriptions of four famous attractions in four countries from the travel review website TripAdvisor, used as sentiment analysis data. This process is relatively simple; due to space limitations, it will not be described here.

(2) Data preprocessing. Section 2.1 introduces the steps involved in online comment text preprocessing.

(3) Keyword extraction and semantic expansion of comment texts. In order to improve classification accuracy, Section 2.2 introduces our online travel review text keyword extraction method and keyword semantic expansion method based on Microsoft Knowledge Graph.

(4) Text representation. Section 2.3 introduces the text representation method based on Word2vec.

(5) Sentiment classification. Section 2.4 introduces the sentiment classification method adopted in this paper.

2.1. Data Preprocessing

Not all characters included in the text of online travel reviews are important. For example, most reviews include words, punctuation, etc. that do not describe the subject of the text. Retaining all characters will lead to the formation of high-dimensional features; this will not only increase the time required for classification learning, but will also introduce a lot of noisy data into the classification and affect the classification accuracy. It is therefore necessary to preprocess the data. The preprocessing process used in this article comprises the following four steps:

1.   Remove HTML tags
2.   Remove non-letters
3.   Convert words to lower case and split them
4.   Remove stop words.

In step 1, Python's BeautifulSoup library was used to remove HTML tags such as '<br>' from the comment text. Steps 2–4 were implemented using NLTK (Natural Language Toolkit) [49] and regular expressions. Here, the second step deletes punctuation, numbers and other non-English characters from the comment text; the third step divides the sentence into words and converts all of these words to lower case; finally, the fourth step uses the stop word list provided by NLTK and deletes these words from the comment text and stoplist. The stop word list contains some noise words that do not describe the text subject ("the", "is", "are", "a", "an", etc.). In addition, combined with the characteristics of the dataset in this article, we added some specific vocabulary words (for example: "Mutianyu", "Great Wall", "China"). These specific high-frequency words will affect the subsequent keyword extraction and sentiment analysis results. However, these words are usually objective descriptions of scenic spots and accordingly do not help with the sentiment analysis.

2.2. Keyword Extraction and Semantic Expansion

The online travel review text obtained in this article pertains to multiple attractions. As shown in Figures 2–5, before preprocessing, the length of the review text about Mutianyu Great Wall, Beijing, China is mostly between 260 and 280 words. Moreover, the length of the comment text for the Harry Potter Wizarding World Theme Park in Orlando, USA is between 90 and 130 words; the comment text for the Tower of London, England is between 90 and 140 words in length; and the lengths of the comment text for the Sydney Opera House in Australia are mostly in two categories (90 to 120 and 200 to 300 words). Because preprocessing will delete some characters that are not related to sentiment classification, the text will be shorter after preprocessing. It is difficult to extract effective feature words from shorter text and thus more difficult to obtain better sentiment classification results [46]. In order to improve the effectiveness of sentiment classification for online travel review text, this paper proposes a keyword semantic expansion method based on knowledge graphs. First, we compared several keyword extraction methods and selected the TextRank method as having the best effect [50] for achieving keyword extraction for online travel review text. Secondly, through the use of Microsoft Knowledge Graph, a conceptual list of keywords for each comment was obtained. This concept list of keywords can be used to expand the semantics of the comment text and provide a richer and more valuable classification feature for the classifier. Next, the specific implementation steps will be introduced.
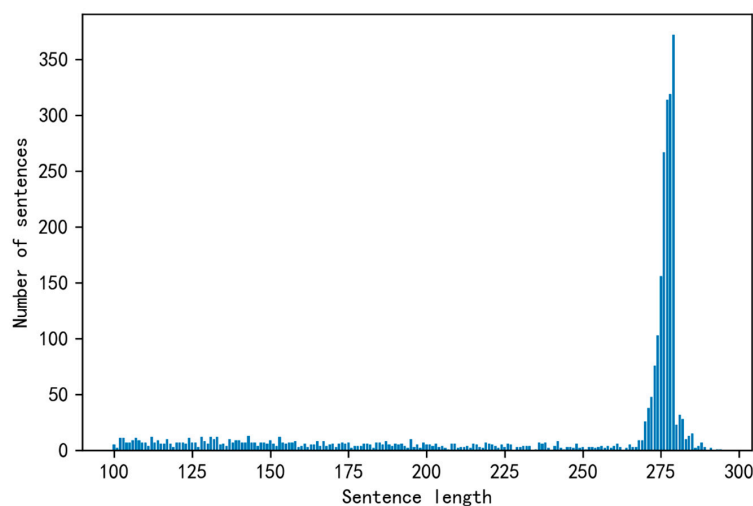

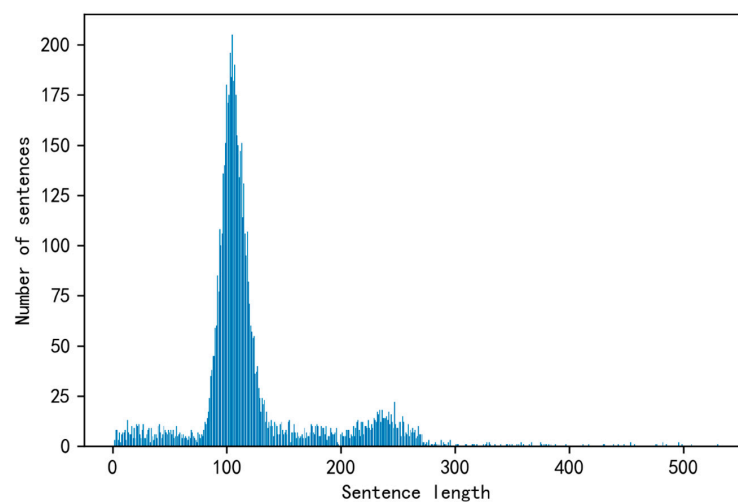**Figure 2.** Length of the comment text of Mutianyu Great Wall in Beijing, China.

Figure 3. Length of the comment text in the Harry Potter Wizarding World Theme Park in Orlando, USA.
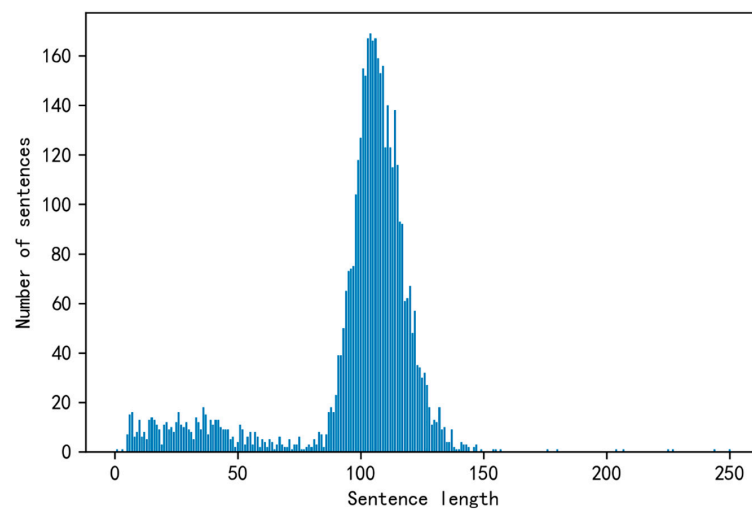
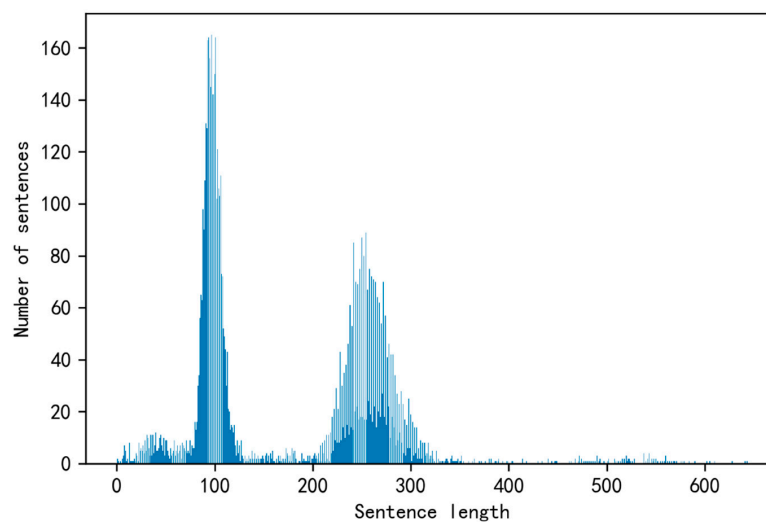**Figure 4.** Length of the comment text of the Tower of London.

Figure 5. Length of the comment text of the Sydney Opera House in Australia.

(1) Keyword extraction

Text keyword extraction is a machine learning algorithm-based text feature extraction method. In fields such as text-based recommendation and search, the accuracy of text keyword extraction is directly related to the final effect. Accordingly, text keyword extraction is an important research direction in the field of text mining. Text keyword extraction methods can be divided into supervised, semi-supervised, and unsupervised methods [51]. Supervised and semi-supervised methods regard keyword extraction as a classification problem and require a labeled training corpus to train the keyword extraction model. However, for massive datasets, labeling the training corpus is often very time-consuming. For its part, the unsupervised keyword extraction method does not require a manually annotated corpus, and is therefore more suitable for the keyword extraction of massive comment texts [52].

The TextRank algorithm proposed by Mihalcea et al. [50] draws on the realization of PageRank, which is the core algorithm of Google search. This is an unsupervised keyword extraction method. Unlike TF-IDF (term frequency—inverse document frequency), LDA (Latent Dirichlet Allocation), etc., TextRank divides the text into several units (e.g., words, sentences) and builds a graph model; keyword extraction can thus be achieved using only the information contained in a single document.

The process by which TextRank extracts text keywords comprises the following steps:

(1) Divide the given text into sentences.

(2) For each sentence segmentation and part-of-speech tagging, filter out stop words, so that only words belonging to the specified part-of-speech are reserved as candidate keywords.

(3) Construct the candidate keyword graph G = (V, E), where V is the node set comprising the candidate keywords generated in step (2); next, use the co-occurrence relationship to construct the edges between any two points.

(4) Calculate the weight of each node. These node weights are sorted in reverse order so that the most important words are obtained as candidate keywords.

(5) Mark the candidate keywords obtained in step (4) in the original text; if adjacent phrases are formed, these are combined into multi-word keywords.

A variety of keyword extraction algorithms represented by TextRank are widely used in tourism and many other fields. Shouzhong et al. [53] integrates TF-IDF and TextRank to mine and analyze personal interests from Weibo text. Paramonov et al. [54] developed a new method combining well-known keyword extraction algorithms (e.g., TextRank and Topic PageRank) and a thesaurus-based procedure, thereby improving the connectivity of the text-via-keyphrase graph while also increasing the accuracy and recall rate of key phrase extraction. Gagliardi et al. [55] integrated the word embedding model and clustering algorithm to establish a novel method capable of automatically extracting keywords/phrases from text without supervision. Ali et al. [56] used the N-gram method to extract the risk factors of heart disease diagnosis and applied these to an intelligent heart disease prediction system, improving the accuracy of heart disease diagnosis.

In Section 3.2, based on the similarity calculation results of the words, and following experiments with TF-IDF and LDA, it is determined that the keywords extracted by TextRank are more suitable for ascertaining the actual semantics of online travel text reviews. Therefore, this study used TextRank for text keyword extraction purposes.

(2) Keyword semantic expansion

Text feature semantic expansion is an effective method of solving the sparse text problem [57]. Wang et al. [58] conceptualized short text into a set of concepts and embedded the original text in order to form word vectors. Experimental results verify that the convolutional neural network based on this word vector can achieve good short text classification results. Rosso et al. [59] believe that combining large-scale unstructured content (text) and high-quality structured data (knowledge graph) can improve text analysis.

Microsoft Knowledge Graph [60] has learned a large amount of common sense knowledge through learning from billions of web pages and years of search logs. The system-provided conceptual model maps text entities into semantic concept categories with specific probabilities; for example, "Microsoft" may automatically map to "software companies" and "Fortune 500 companies" [61]. This paper introduces the conceptual model of the Microsoft Knowledge Graph to expand the semantics of online travel review text keywords. This knowledge graph-based keyword semantic expansion method utilizes the huge information corpus of the Microsoft knowledge graph to expand the semantics of the text. This method overcomes the issue of fewer features being available that is caused by the sparseness of short texts, and accordingly provides richer and more valuable classification features for short text sentiment classification. We demonstrate the improvement in classification accuracy brought about by this method in the experiment discussed in Chapter 3.

2.3. Text Representation

(1) Text representation of comments based on Word2vec

Representing text as structured data that is able to be handled by machine learning classification algorithms is a highly important part of the text classification process. In 2013, Google released the software tool Word2vec for training word vectors [62]. Word2vec's high-dimensional vector model solves the multi-dimensional semantic problem, because it can quickly and effectively express words in high-dimensional vector form through the optimized training model according to a given corpus,

thereby providing a new tool for the application research in the field of natural language processing [63]. Academic research [64,65] demonstrates that Word2vec has achieved excellent performance in the fields of text similarity calculation and text classification. In light of the above analysis, this study opted to construct Word2vec vectors for the pre-processed and semantically expanded comment text.

(2) Data normalization

Normalized data exhibits enhance stability for attributes with very small variances, while maintaining 0 entries in the sparse matrix [62]. Therefore, this study used the normalization method to scale the text vector represented by Word2vec to between 1 and 0. The formula utilized is as follows:

$$x_i' = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{1}$$

In Equation (1), x_i' represents the result of normalization, while x_i represents the data that needs to be normalized. Moreover, x_{max} and x_{min} represent the maximum and minimum values in the dataset, respectively.

2.4. Sentiment Classification

For massive texts, one effective solution involves transforming sentiment analysis into classification and applying machine learning methods in order to solve such problems [66]. This article has introduced the problems encountered by deep learning methods, along with the excellent results achieved by machine learning methods in the text sentiment analysis context. Therefore, using the online travel review text data processed in Sections 2.1–2.3 as the training data, this SVM was chosen in this study as the method of sentiment classification. In Section 3.4, through the analysis of experimental results, the most suitable sentiment classification model for processing online travel review texts is then provided.

3. Case Study

This section introduces the research process utilized in this article and draws conclusions from a sentiment classification experiment on online tourist review texts of multiple attractions. In more detail, Section 3.1 describes the experimental dataset and the results of preprocessing; Section 3.2 introduces the experimental process of keyword semantic expansion based on the knowledge graph; Section 3.3 introduces the text representation based on Word2vec; finally, Section 3.4 introduces the sentiment classification based on SVM experiments and result analysis.

3.1. Data Acquisition and Preprocessing Experiment

As shown in Table 2, the present research used a crawler program written in Python to obtain four datasets from the TripAdvisor website. Mutianyu_Great_Wall contains review text pertaining to Mutianyu Great Wall, which is the number one attraction on the TripAdvisor website in Beijing, China. It contains a total of 2772 pieces of review data in English published by tourists from January 2016 to December 2019. Wizarding_World_of_Harry_Potter contains comment text pertaining to the Harry Potter Wizarding World Theme Park in Orlando, USA, specifically, 6641 pieces of comment data published by tourists from June 2017 to December 2019 in English. Tower_of_London contains comment text about the Tower of London in the United Kingdom. It contains the data of 4428 comments in English published by tourists from July 2018 to December 2019. Finally, Sydney_Opera_House contains review text pertaining to the Sydney Opera House in Australia. It contains 6776 pieces of comment data in English published by tourists from March 2017 to December 2019.

Table 3 presents a piece of comment text in the Mutianyu_Great_Wall dataset and its pre-processed results. In the table, the first column is the comment text published by tourists, while the second column is the pre-processed text. As can be seen from the introduction in Section 2.2, the preprocessing operation removes any original comment text content that is unrelated to sentiment classification.

Table 2. Experimental data set.

Dataset	URL	Size
Mutianyu_Great_Wall	https://www.tripadvisor.cn/Attraction_Review-g294212-d325811-Reviews-Mutianyu_Great_Wall-Beijing.html	2772
Wizarding_World_of_Harry_Potter	https://www.tripadvisor.cn/Attraction_Review-g34515-d1968468-Reviews-The_Wizarding_World_of_Harry_Potter-Orlando_Florida.html	6641
Tower_of_London	https://www.tripadvisor.cn/Attraction_Review-g186338-d187547-Reviews-Tower_of_London-London_England.html	4428
Sydney_Opera_House	https://www.tripadvisor.cn/Attraction_Review-g255060-d257278-Reviews-Sydney_Opera_House-Sydney_New_South_Wales.html	6776

Table 3. Review and preprocessed results.

Review	Preprocessed Review
If you want to go individually, not with organized tour, we were offered the option to go by taxi. The taxi left us to the bus station we bought tickets for the bus, the entrance and the cable. The bus left us at the cable station and the cable took us to the Wall. It was very easy and not so expensive and we arranged for the hours which were convenient for us.	individually organized tour offered option taxi taxi left bus station bought tickets bus entrance cable bus left cable station cable easy expensive arranged hours convenient

3.2. Keyword Semantic Expansion Experiment Based on Knowledge Graph

(1) Comparison of text keyword extraction methods

In this study, three types of text keyword extraction methods, namely TF-IDF, LDA, and TextRank were selected to carry out comparative experiments. Taking the comment text in Table 3 as an example, the manually provided subject term is "transport", while the second column of Table 4 presents the keywords with the largest calculation result values obtained by each of the three methods. Among them, the calculation results of the two keywords obtained by TF-IDF are the same.

Table 4. Results of three types of keyword extraction methods.

Method	Keyword	Similarity Result
TF-IDF	cable, bus	cable: 0.139 bus: 0.379
LDA	offered	offered: −0.022
TextRank	bus	bus: 0.379

Word2vec is able to convert words into vectors and calculate the distance between the vectors. The larger the value of the calculation result, the greater the similarity between the two words [63]. Based on Word2vec's word vector similarity calculation, this study calculated the similarity of the word vector using the first-order keywords obtained by the three methods in addition to the subject words ("transport") of the manually provided comment text. The calculation results are shown in the third column of Table 4. Here, the keywords obtained by TextRank are the most relevant to the subject words of the manually provided review text. TF-IDF also identified the most relevant keyword, "bus". However, the keyword "cable", which has the same weight as "bus", has poor relevance to the subject words of the manually provided review text, which affects the final result. LDA requires a large corpus (i.e., large amount of comment text) for accurate results to be obtained. However, this research requires keywords to be derived for each short text of the comment. Therefore, LDA is unsuitable for this research, and the final keyword extraction effect is also poor. This study randomly selected 10% of the samples in each dataset and used the above three methods to extract keywords.

Following experimental comparison, TextRank was found to have the best keyword extraction effect. Therefore, TextRank was used to extract the text keywords from online travel reviews.

(2) Keyword semantic expansion experiment

This study obtained a concept list of online travel review text keywords using the conceptual model of Microsoft Knowledge Graph [67]. For example, the conceptual list of the keyword "bus" in the comment text of Table 3 is as follows: vehicle, public transportation, large vehicle, etc.

For the four datasets listed in Table 2, TextRank was used to extract text keywords from a total of 20,617 comment texts. In the next step, the concept list for the first-ranked keywords was obtained in ascending order of weight. Although Microsoft Knowledge Graph covers a very wide range, it does not cover any word. Following calculation, Microsoft Knowledge Graph returned results for 97.6% of the keywords in this experiment. Finally, we added the return results of each keyword to the pre-processed comment text in order to create a pre-classification dataset.

3.3. Text Representation and Normalization Experiment

Once preprocessing and semantic expansion was complete, the comment text was typically under 300 characters in length. Therefore, googlenews-vecctors-negative300.bin [62], a word vector library of news corpora pre-trained by Google, was selected to create a comment text vector. The final results are illustrated in Figure 6. Each line in the figure is a normalized 300-dimensional Word2vec real vector, which represents a specific comment text.

	0	1	2	3	4	5	6	7	8	9	...	290	291	292	293	294	295	296	297	298	299
0	-0.033	0.050	0.026	-0.030	-0.003	-0.040	0.084	-0.133	0.134	0.143	...	-0.093	0.083	-0.087	-0.061	-0.034	-0.097	-0.007	-0.057	0.059	-0.061
1	0.055	-0.092	0.033	-0.023	-0.072	-0.088	0.115	-0.009	0.173	0.011	...	-0.088	0.023	-0.119	0.013	-0.100	-0.098	-0.021	0.039	0.017	-0.099
2	0.009	-0.018	-0.072	0.078	-0.019	0.053	0.172	0.044	0.090	0.006	...	-0.027	0.050	-0.084	-0.044	-0.048	-0.014	-0.022	0.006	0.052	-0.048
3	0.017	-0.005	-0.006	0.043	-0.041	-0.120	0.060	-0.110	0.096	0.037	...	-0.065	-0.000	-0.135	-0.043	-0.036	-0.106	-0.012	-0.025	0.070	-0.001
4	0.006	0.077	-0.037	0.019	-0.078	-0.032	0.087	-0.058	0.078	0.098	...	-0.035	0.036	-0.049	0.067	-0.077	-0.012	-0.051	-0.028	0.059	0.025

Figure 6. Vector representation of text.

3.4. Sentiment Classification Experiment

(1) Acquisition of training set classification labels

The machine learning classification method represented by SVM requires training data with sentiment classification results for model training. The sentiment classification results for these training data are also referred to as the training set classification labels. In this study, manual analysis and sentiment analysis software were used to generate the classification labels for the training set.

SentiStrength [68] is a software package that estimates the strength of positive and negative emotions contained in text. It also has an artificial level of accuracy for short social network texts in English. We chose the nine-level sentiment classification results provided by SentiStrength. For negative emotions, the scores range from -1 (not negative) to -4 (extremely negative); for positive emotions, moreover, the scores range from 1 (not positive) to 4 (extremely positive); 0 represents neutral emotion. In this study, the SentiStrength results were again scored by humans, and the adjustment rate was about 24.7%. Finally, the sentiment analysis results of the dataset are presented in Figures 7–10. The abscissa represents the sentiment analysis results, which range from -4 to 4 in a total of nine categories; the ordinate indicates the number of samples in each category. It can be seen that the number of samples of each sentiment value is extremely uneven. For example, in the review dataset for the Mutianyu Great Wall in Beijing, China, the number of texts in category 2 is 1266, while there is only 1 text in category -4. Moreover, there is no category -4 data in the review data of the Sydney Opera House in Australia.
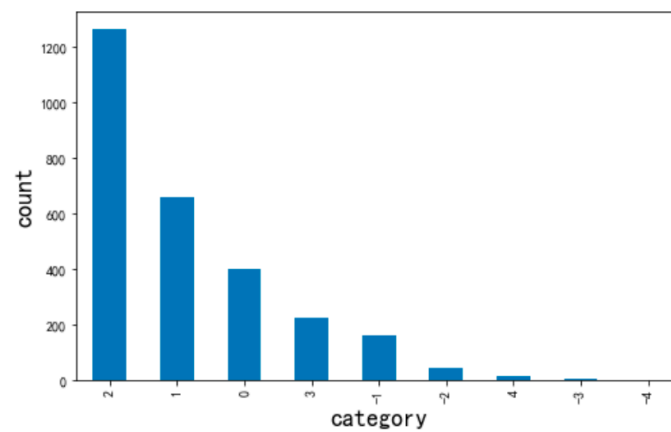
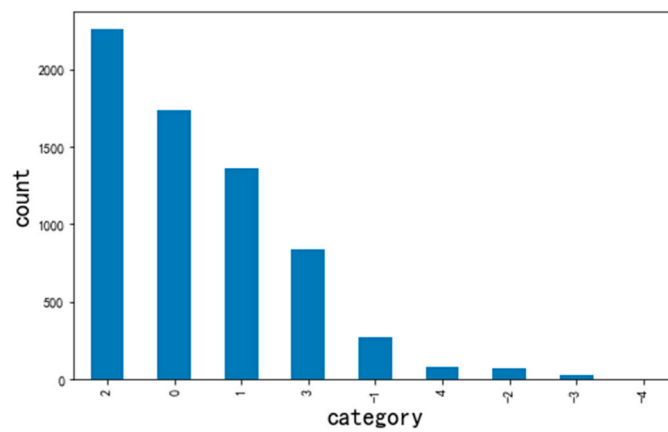**Figure 7.** The sentiment distribution of the review text of Mutianyu Great Wall in Beijing, China.

Figure 8. The sentiment distribution of the review text of the Harry Potter Wizarding World Theme Park in Orlando, USA.
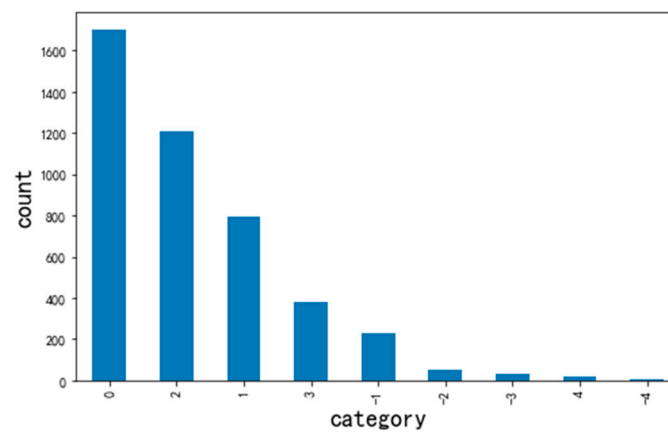

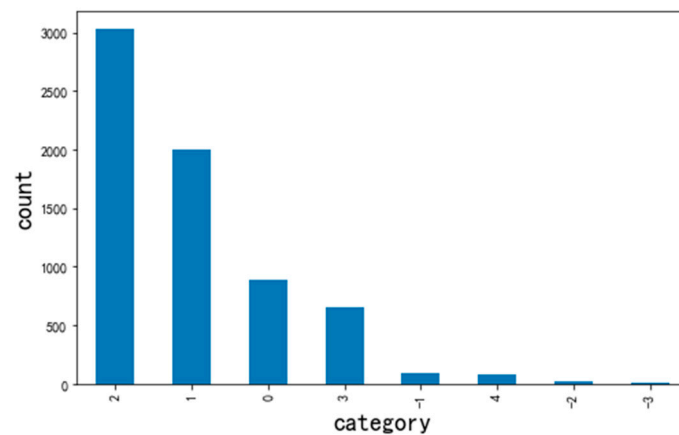**Figure 9.** The sentiment distribution of the review text of Tower of London.

Figure 10. The sentiment distribution of the review text of Sydney Opera House, Australia.

(2) Sampling experiment of unbalanced data

From the analysis presented in the previous section, we can see that the sentiment distribution of online travel review texts is very uneven. In fact, this is a typical unbalanced dataset. For unbalanced datasets, machine learning classifiers will tend to incorrectly divide new samples into categories with more samples, resulting in classification errors [69]. The methods used to process unbalanced datasets are mainly divided into undersampling, oversampling, and improved methods [70]. This study used Python to implement two types of sampling methods. Our experimental results demonstrate that, due to the extremely uneven sentiment distribution of the experimental dataset used, the undersampling dataset was so small that it was difficult to obtain more accurate classification results. Overall, Naive Random Over Sampler (ROS) [71] achieved the best sampling results.

(3) Evaluation index

The evaluation indicators of classification results that have been adopted by academia include Accuracy, Precision, Recall, and F1 score [72]. In binary classification, the sample categories are divided into positive and negative types. Let us suppose that TP represents the number of samples that are both actually positive and classified as positive, while FP denotes the number of samples that are actually negative but classified as positive; moreover, FN represents the number of samples that are in fact positive but are classified as negative, while TN indicates the number of samples that are both actually negative and classified as negative. In addition, the accuracy rate refers to the proportion of correct samples classified as positive to the samples classified as positive. The calculation formula for this is as follows:

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

Furthermore, the recall rate refers to the proportion of correct samples classified as positive to actually positive samples, and the calculation formula is as follows:

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

Finally, the F1 score is the harmonic average of the precision rate and the recall rate. The calculation formula is as follows:

$$F1\ score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{4}$$

The accuracy rate reflects the model's ability to distinguish negative samples: the higher the accuracy rate, the stronger the model's ability to distinguish negative samples. Moreover, the recall rate reflects the model's ability to identify positive samples: the higher the recall rate, the stronger the model's ability to recognize positive samples. In addition, the F1 score is the combination of the accuracy rate and recall rate: the higher the F1 score, the more robust the model. While accuracy is the

simplest and most intuitive evaluation index in classification, it is also affected by obvious defects. For example, if we assume that 99% of the samples are positive samples, the classifier could obtain 99% accuracy if it always predicted a positive result, but its actual performance would be very low. That is to say, when the proportion of samples in different categories is highly uneven, the category with the largest proportion often becomes the most important factor affecting the accuracy. As the experimental data in this study was unbalanced data, we did not use the accuracy rate as a classification result evaluation index. Instead, we selected three indicators—accuracy rate, recall rate, and F1 score—to measure the classification results.

(4) SVM-based sentiment classification

Python's sklearn was used to implement the SVM algorithm. After a large number of experiments, the kernel function RBF (Radial Basis Function) was found to achieve the highest classification accuracy, while other parameters were assigned default values. We used 30% of the data as test data and the remaining 70% as training data.

The comparative experimental results of one dataset (Mutianyu_Great_Wall) are presented in Table 5. The first row of Table 5 displays the classification results of SVM. The classification accuracy of SVM on this imbalanced dataset is very low, as it assigns most of the samples to the category with the largest number of samples. Once ROS sampling and the Word2vec vectorization of text is complete, the data in the second row of Table 5 shows that the SVM algorithm's classification result has been greatly improved. The next experiment carried out involves extracting TextRank keywords from the comment text and expand the semantics of the keywords with the largest weights based on the Microsoft Knowledge Graph. The semantic expansion of keywords and pre-processed online travel review text make up the SVM classification dataset. Moreover, the experimental results in the third row of Table 5 list the final classification results; it can be seen from this table that the knowledge graph-based keyword semantic expansion method proposed in this paper optimizes the classification results.

Table 5. Mutianyu_Great_Wall experiment results.

Method	Precision	Recall	F1 score
SVM	0.108	0.238	0.148
SVM + ROS + Word2vec	0.840	0.822	0.830
SVM + ROS + Word2vec + KnowledgeGraph	**0.859**	**0.901**	**0.879**

Optimal solution in the comparison result is marked in bold.

Table 6 presents the experimental results of the other three datasets. Similar to the experimental results of the Mutianyu_Great_Wall dataset, it can be seen that the sampling technique, Word2vec-based text vectorization, and knowledge graph-based keyword semantic expansion method effectively improve the classification effect. Similar experimental results obtained on different datasets verify the universality of this method. In short, this provides an effective solution for sentiment analysis of online travel review texts.

Table 6. Experimental results of the other three data sets.

Method	Wizarding_World_of_Harry_Potter			Tower_of_London			Sydney_Opera_House		
	Precision	Recall	F1 Score	Precision	Recall	F1 Score	Precision	Recall	F1 Score
SVM	0.108	0.339	0.164	0.701	0.404	0.512	0.287	0.445	0.348
SVM + ROS + Word2vec	0.637	0.643	0.639	0.833	0.868	0.85	0.702	0.716	0.708
SVM + ROS + Word2vec + KnowledgeGraph	**0.823**	**0.843**	**0.832**	**0.933**	**0.944**	**0.938**	**0.820**	**0.837**	**0.828**

Optimal solution in the comparison result is marked in bold.

The receiver operating characteristic curve (ROC) is an evaluation method that demonstrates the accuracy of classification through intuitive graphics. Figures 11–14 show the ROC curves of the four data sets. We have labeled each sentiment category (the Sydney_Opera_House dataset has only eight sentiment categories) with a different color. The abscissa in the figures indicates the proportion of samples classified as positive but actually negative to all negative samples; the ordinate represents the proportion of all positive samples that are predicted to be positive and actually positive. The closer the ROC curve is to the upper left corner, the higher the accuracy of the experiment.
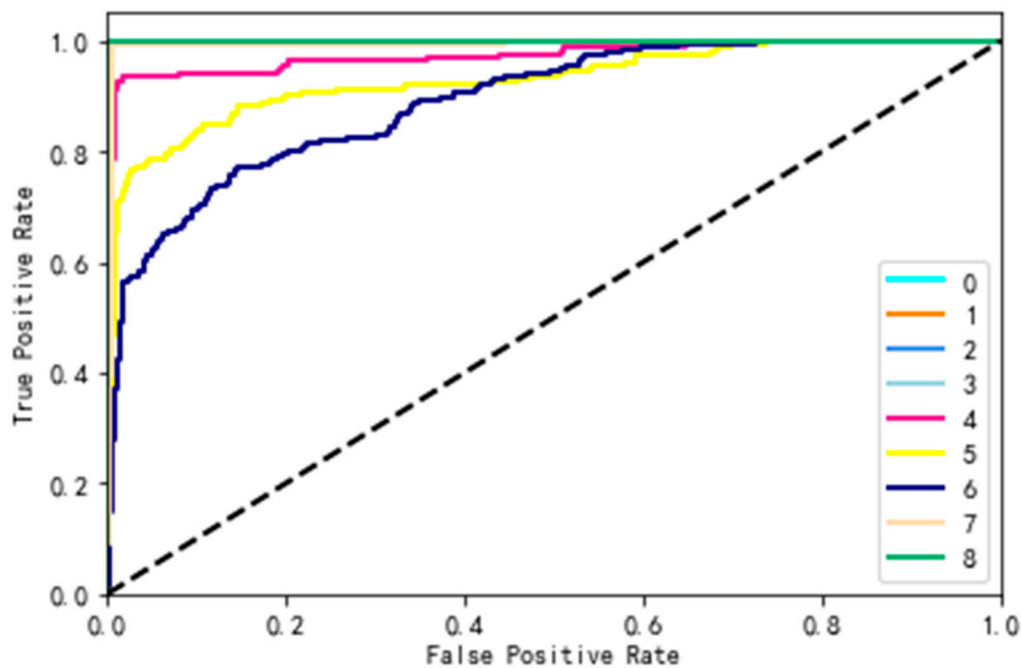

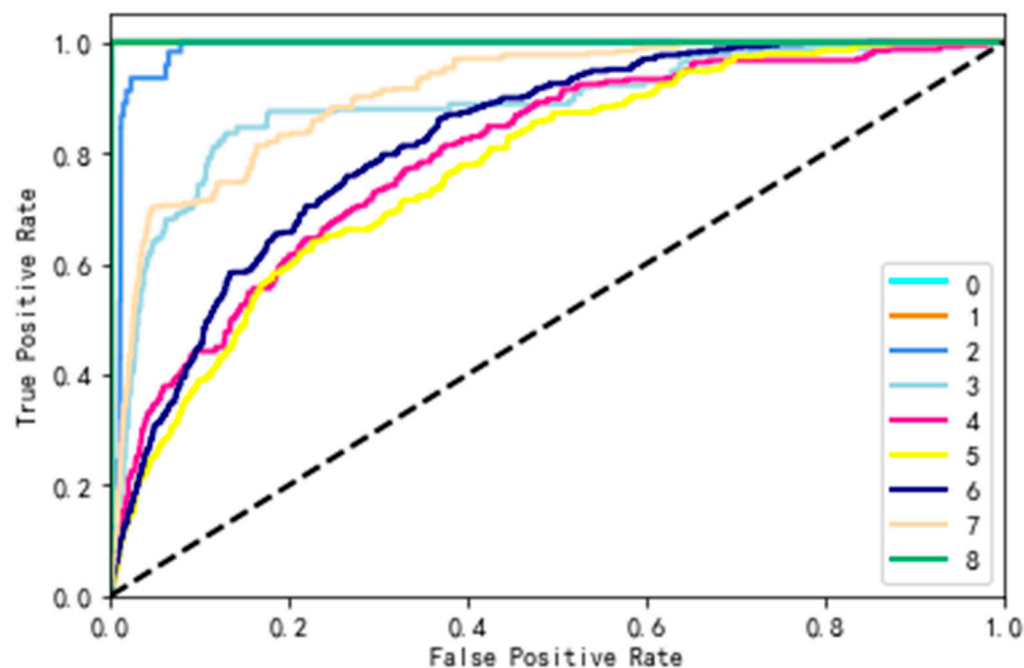**Figure 11.** Mutianyu_Great_Wall receiver operating characteristic curve.

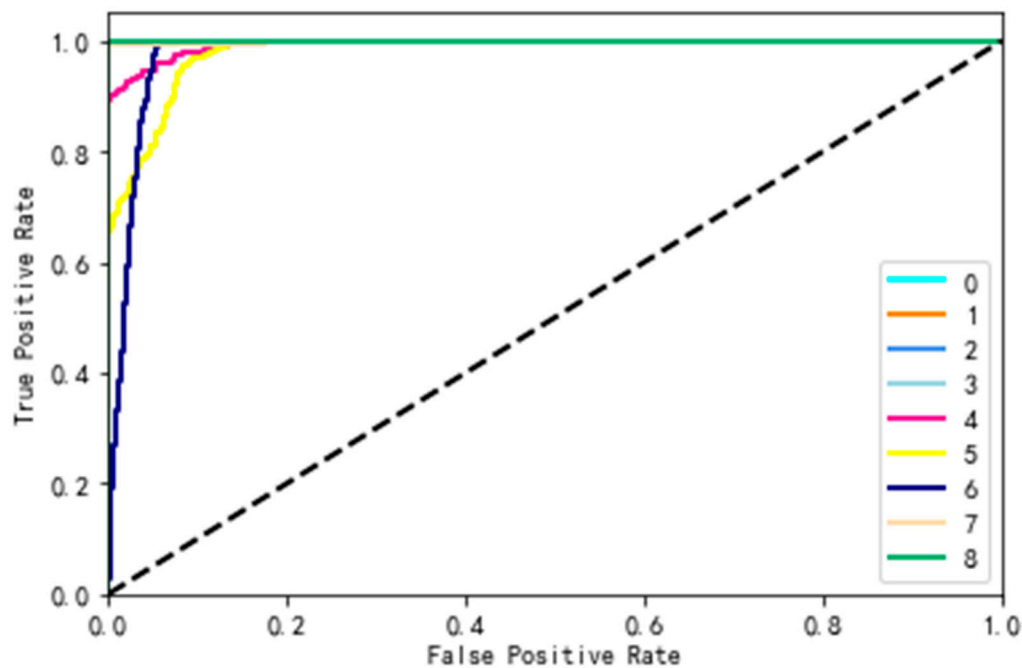**Figure 12.** Wizarding_World_of_Harry_Potter receiver operating characteristic curve.

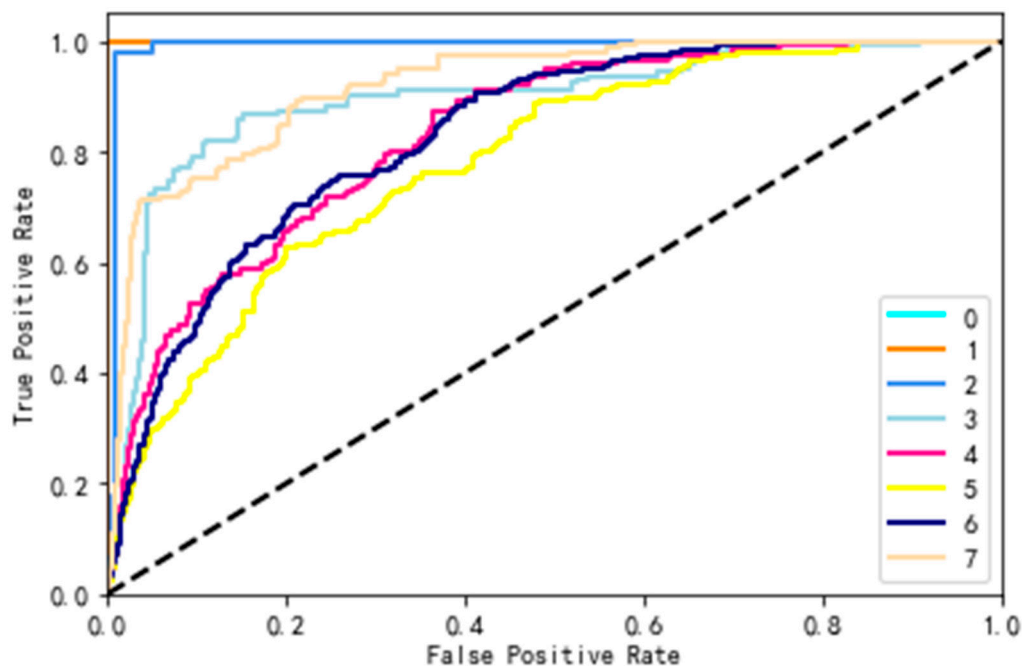**Figure 13.** Tower_of_London receiver operating characteristic curve.

Figure 14. Sydney_Opera_House receiver operating characteristic curve.

4. Discussion

Sentiment analysis is a mainstream technology that employs social media analysis strategies to analyze customer feedback and comments. Conducting sentiment analysis based on websites such as TripAdvisor is desirable because a large number of free datasets can be obtained from such websites for large-scale research, while such large-scale data cannot easily be obtained via traditional research methods. Big data provides a new type of data for use in tourism research, and also puts forward higher requirements for data processing. Currently, few studies have been conducted on the applicability and accuracy of sentiment analysis methods in the tourism research literature. In addition,

contemporary research ignores the possibility of integrating human knowledge, such as knowledge graphs, into existing methods in order to improve the text sentiment analysis performance. Big data is characterized by a huge data volume, and the speed and accuracy requirements for sentiment analysis are becoming steadily higher [8]. Therefore, the prospect of developing suitable and efficient sentiment analysis methods for specific types of big data in the tourism context is a highly valuable proposition.

The obtained sentiment analysis results based on TripAdvisor review text can be applied to multiple fields. For example, they can help sightseeing spots, restaurants, or hotels to explain comments and adopt corresponding countermeasures, which can in turn provide decision makers and customers with better decision-making information. Similarly, this approach can also be used to study theoretical issues related to customer satisfaction (for example, whether a tour guide service would improve the tourist experience). However, existing studies [43,44] have found that the key features of the review text differ substantially depending on which websites they are drawn from, and that it is therefore necessary to conduct sentiment analysis research on one specific website at a time. Therefore, research into machine learning sentiment analysis methods for TripAdvisor review texts will aid in the development of tourism research utilizing these texts. Compared with vocabulary sentiment analysis, one of the advantages of machine learning sentiment analysis is that it does not require humans to create a dictionary; this is beneficial because the production of such a dictionary is a time-consuming and laborious process. In addition, machine learning methods achieve more accurate performance on larger amounts of training data than can be obtained using vocabulary sentiment analysis [8]. Feature extraction is a key issue in the application of machine learning to the field of sentiment analysis [24]. Accordingly, this study designed and implemented a sentiment classification method based on the semantic expansion of text keywords that both increases the classification features and improves the accuracy of sentiment analysis, thereby providing a novel solution for machine learning sentiment analysis.

In terms of the specific details of the work of this article, in order to improve the accuracy of sentiment analysis conducted on online travel review texts, this study conducted extensive research work on the classification problems caused by the data features of online review texts. First, most online review texts are short texts, which makes it difficult to obtain more accurate sentiment classification results. To solve this problem, we designed a text keyword semantic expansion method based on a knowledge graph. In this part of the research, the present study compared three typical text keyword extraction methods and provided keyword extraction methods that are suitable for online travel review texts. In addition, based on Microsoft Knowledge Graph, the semantics of text keywords were expanded, and richer and more valuable sentiment classification features were constructed. The second part of the research involved comparing the two types of sampling methods and identifying which of these is more suitable for use in solving the uneven sentiment distribution problem in online review texts. This article fully describes the key aspects of online travel review text sentiment classification, establishes an effective sentiment classification research framework for online travel review text, and validates the proposed method based on a relatively extensive sample.

The work put forward in this paper aims to emphasize and improve the methodological relevance and applicability of sentiment analysis. However, there are some limitations:

- Studies have shown [10] that deleting comment text without emotional content can improve the accuracy of emotional classification. This idea is worth examining in future.
- In terms of keyword selection, this study only selected the keyword with the largest TextRank value. The questions of how to choose keywords with the same value, whether more keywords can be introduced for semantic expansion, the relationship between these choices, and the accuracy of sentiment classification are also worthy of further study.
- In the Word2vec-based text representation method, the use of different Word2vec corpora will yield different results. The best approach would be to train a corpus of specific topics [73].
- In terms of automated classification methods, studies have shown that the combination of LSTM (Long Short-Term Memory) and attention mechanisms [74] has resulted in excellent emotion

classification results. However, the question of whether these novel methods are suitable for the research object of this article is worthy of further study.

- In terms of experimental subjects, this article only studies English reviews from TripAdvisor, and does not investigate other online travel platforms and other languages. Therefore, it is highly advisable to investigate data in other languages and other platforms to verify the applicability of this method.

Author Contributions: Conceptualization, W.C., Z.X. and Y.L.; Methodology, W.C.; Software, W.C.; Validation, W.C., Q.Y. and X.Z.; Data curation, W.C.; Writing—original draft preparation, W.C.; Writing—review and editing, W.C., Z.X. and Y.L.; Supervision, Y.L.; Project administration, Y.L.; Funding acquisition X.Z., Q.Y. and Y.L. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Li, X.; Pan, B.; Law, R.; Huang, X. Forecasting tourism demand with composite search index. *Tour. Manag.* **2017**, *59*, 57–66. [CrossRef]
2. Sivarajah, U.; Kamal, M.M.; Irani, Z.; Weerakkody, V. Critical analysis of Big Data challenges and analytical methods. *J. Bus. Res* **2017**, *70*, 263–286. [CrossRef]
3. Guo, Y.; Barnes, S.J.; Jia, Q. Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tour. Manag.* **2017**, *59*, 467–483. [CrossRef]
4. Chuang, S. Co-creating social media agility to build strong customer-firm relationships. *Ind. Mark. Manag.* **2020**, *84*, 202–211. [CrossRef]
5. Kauffmann, E.; Peral, J.; Gil, D.; Ferrandez, A.; Sellers, R.; Mora, H. A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making. *Ind. Mark. Manag.* **2019**, in press. [CrossRef]
6. Li, J.; Xu, L.; Tang, L.; Wang, S.; Li, L. Big data in tourism research: A literature review. *Tour. Manag.* **2018**, *68*, 301–323. [CrossRef]
7. Fang, B.; Ye, Q.; Kucukusta, D.; Law, R. Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics. *Tour. Manag.* **2016**, *52*, 498–506. [CrossRef]
8. Alaei, A.; Becken, S.; Stantic, B. Sentiment Analysis in Tourism: Capitalizing on Big Data. *J. Travel Res.* **2019**, *58*, 175–191. [CrossRef]
9. Asghar, M.Z.; Kundi, F.M.; Ahmad, S.; Khan, A.; Khan, F. T-SAF: Twitter sentiment analysis framework using a hybrid classification scheme. *Expert Syst.* **2018**, *35*, e12233. [CrossRef]
10. Afzaal, M.; Usman, M.; Fong, A.; Fong, S. Multiaspect-based opinion classification model for tourist reviews. *Expert Syst.* **2019**, *36*, e12371. [CrossRef]
11. Gunther, W.; Mehrizi, M.H.R.; Huysman, M.; Feldberg, F. Debating big data: A literature review on realizing value from big data. *J. Strateg. Inf. Syst.* **2017**, *26*, 191–209. [CrossRef]
12. Mariani, M.M.; Baggio, R.; Fuchs, M.; Hoepken, W. Business intelligence and big data in hospitality and tourism: A systematic literature review. *Int. J. Contemp. Hosp. Manag.* **2018**, *30*, 3514–3554. [CrossRef]
13. Kirilenko, A.P.; Stepchenkova, S.; Kim, H.; Li, X. Automated Sentiment Analysis in Tourism: Comparison of Approaches. *J. Travel Res.* **2018**, *57*, 1012–1025. [CrossRef]
14. Ali, F.; Kwak, D.; Khan, P.; Islam, S.M.R.; Kim, K.; Kwak, K.S. Fuzzy ontology-based sentiment analysis of transportation and city feature reviews for safe traveling. *Transp. Res. Part C Emerg.* **2017**, *77*, 33–48. [CrossRef]
15. Luo, Y.; Tang, R. Understanding hidden dimensions in textual reviews on Airbnb: An application of modified latent aspect rating analysis (LARA). *Int. J. Hosp. Manag.* **2019**, *80*, 144–154. [CrossRef]
16. Pang, B.; Lee, L. Opinion Mining and Sentiment Analysis. *Found. Trends Inf. Retr.* **2008**, *2*, 1–135. [CrossRef]
17. Neidhardt, J.; Rummele, N.; Werthner, H. Can We Predict Your Sentiments by Listening to Your Peers. In *Information and Communication Technologies in Tourism 2016*; Springer: Cham, Switzerland, 2016; pp. 593–603.

18. Garciapablos, A.; Cuadros, M.; Linaza, M.T. Automatic analysis of textual hotel reviews. *Inf. Technol. Tour.* **2016**, *16*, 45–69. [CrossRef]

19. An, H.; Moon, N. Design of recommendation system for tourist spot using sentiment analysis based on CNN-LSTM. *J. Ambient Intell. Humaniz. Comput.* **2019**, 1–11. [CrossRef]

20. Pan, D.; Yuan, J.; Li, L.; Sheng, D. Deep neural network-based classification model for Sentiment Analysis. *arXiv* **2019**, arXiv:1907.02046.

21. Handhika, T.; Fahrurozi, A.; Sari, I.; Lestari, D.P.; Zen, R.I. Hybrid Method for Sentiment Analysis Using Homogeneous Ensemble Classifier. In Proceedings of the 2019 2nd International Conference of Computer and Informatics Engineering (IC2IE), Banyuwangi, Indonesia, 10–11 September 2019; pp. 232–236.

22. Kim, K.; Park, O.; Yun, S.; Yun, H. What makes tourists feel negatively about tourism destinations? Application of hybrid text mining methodology to smart destination management. *Technol. Forecast. Soc. Chang.* **2017**, *123*, 362–369. [CrossRef]

23. Hinterstoisser, S.; Lepetit, V.; Wohlhart, P.; Konolige, K. On Pre-Trained Image Features and Synthetic Images for Deep Learning. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.

24. Ghiassi, M.; Lee, S. A domain transferable lexicon set for Twitter sentiment analysis using a supervised machine learning approach. *Expert Syst. Appl.* **2018**, *106*, 197–216. [CrossRef]

25. Kim, K. An improved semi-supervised dimensionality reduction using feature weighting: Application to sentiment analysis. *Expert Syst. Appl.* **2018**, *109*, 49–65. [CrossRef]

26. Ali, F.; Kwak, K.-S.; Kim, Y.-G. Opinion mining based on fuzzy domain ontology and Support Vector Machine: A proposal to automate online review classification. *Appl. Soft Comput.* **2016**, *47*, 235–250. [CrossRef]

27. Parlar, T.; Özel, S.A.; Song, F. QER: A new feature selection method for sentiment analysis. *Hum. Cent. Comput. Inf. Sci.* **2018**, *8*, 10. [CrossRef]

28. Zainuddin, N.; Selamat, A.; Ibrahim, R. Hybrid sentiment classification on twitter aspect-based sentiment analysis. *Appl. Intell.* **2018**, *48*, 1218–1232. [CrossRef]

29. Kumar, A.; Jaiswal, A. Swarm intelligence based optimal feature selection for enhanced predictive sentiment accuracy on twitter. *Multimed. Tools Appl.* **2019**, *78*, 29529–29553. [CrossRef]

30. Pu, X.; Wu, G.; Yuan, C. Exploring overall opinions for document level sentiment classification with structural SVM. *Multimed. Syst.* **2019**, *25*, 21–33. [CrossRef]

31. Adhi, M.S.; Nafan, M.Z.; Usada, E. Pengaruh Semantic Expansion pada Naïve Bayes Classifier untuk Analisis Sentimen Tokoh Masyarakat. *J. RESTI* **2019**, *3*, 141–147. [CrossRef]

32. Fang, C.; Huang, Y. Knowledge-enhanced neural networks for sentiment analysis of Chinese reviews. *Neurocomputing* **2019**, *368*, 51–58.

33. Alowaidi, S.; Saleh, M.; Abulnaja, O. Semantic Sentiment Analysis of Arabic Texts. *Int. J. Adv. Comput. Sci. Appl.* **2017**, *8*, 256–262. [CrossRef]

34. Asgarian, E.; Kahani, M.; Sharifi, S. The Impact of Sentiment Features on the Sentiment Polarity Classification in Persian Reviews. *Cogn. Comput.* **2018**, *10*, 117–135. [CrossRef]

35. Agarwal, B.; Mittal, N. Sentiment Analysis Using ConceptNet Ontology and Context Information. In *Prominent Feature Extraction for Sentiment Analysis*; Springer: Cham, Switzerland, 2016; pp. 63–75.

36. Höpken, W.; Fuchs, M.; Menner, T.; Lexhagen, M. Sensing the Online Social Sphere Using a Sentiment Analytical Approach. In *Analytics in Smart Tourism Design: Concepts and Methods*; Xiang, Z., Fesenmaier, D.R., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 129–146. ISBN 978-3-319-44263-1.

37. Akhtar, N.; Zubair, N.; Kumar, A.; Ahmad, T. Aspect based Sentiment Oriented Summarization of Hotel Reviews. *Procedia Comput. Sci.* **2017**, *115*, 563–571. [CrossRef]

38. Ma, E.; Cheng, M.; Hsiao, A. Sentiment analysis – a review and agenda for future research in hospitality contexts. *Int. J. Contemp. Hosp. Manag.* **2018**, *30*, 3287–3308. [CrossRef]

39. Ko, C. Exploring Big Data Applied in the Hotel Guest Experience. *Open Access Libr. J.* **2018**, *5*, 1–17. [CrossRef]

40. Stepchenkova, S.; Kirilenko, A.P.; Li, X. Barriers and Sentiment of the American Tourists Toward Travel to China. In *Tourist Behavior*; Springer: Cham, Switzerland, 2018; pp. 129–139.

41. Bansal, B.; Srivastava, S. Hybrid attribute based sentiment classification of online reviews for consumer intelligence. *Appl. Intell.* **2019**, *49*, 137–149. [CrossRef]

42. Lawani, A.; Reed, M.R.; Mark, T.B.; Zheng, Y. Reviews and Price on Online Platforms: Evidence from Sentiment analysis of Airbnb reviews in Boston. *Reg. Sci. Urban Econ.* **2019**, *75*, 22–34. [CrossRef]

43. Valdivia, A.; Luzón, M.V.; Herrera, F. Sentiment Analysis in TripAdvisor. *IEEE Intell. Syst.* **2017**, *32*, 72–77. [CrossRef]

44. Xiang, Z.; Du, Q.; Ma, Y.; Fan, W. A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tour. Manag.* **2017**, *58*, 51–65. [CrossRef]

45. Mirzaalian, F.; Halpenny, E. Social media analytics in hospitality and tourism: A systematic literature review and future trends. *J. Hosp. Tour. Technol.* **2019**, *10*, 764–790. [CrossRef]

46. Zhang, W.; Kong, S.; Zhu, Y.; Wang, X. Sentiment classification and computing for online reviews by a hybrid SVM and LSA based approach. *Clust. Comput.* **2019**, *22*, 12619–12632. [CrossRef]

47. Valdivia, A.; Hrabova, E.; Chaturvedi, I.; Luzon, M.V.; Troiano, L.; Cambria, E.; Herrera, F. Inconsistencies on TripAdvisor reviews: A unified index between users and Sentiment Analysis Methods. *Neurocomputing* **2019**, *353*, 3–16. [CrossRef]

48. Schmunk, S.; Hopken, W.; Fuchs, M.; Lexhagen, M. Sentiment Analysis: Extracting Decision-Relevant Knowledge from UGC. In *Information and Communication Technologies in Tourism 2014*; Springer: Cham, Switzerland, 2013; pp. 253–265.

49. Natural Language Toolkit. Available online: http://www.nltk.org/ (accessed on 17 June 2020).

50. Mihalcea, R.; Tarau, P. TextRank: Bringing Order into Texts. In Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, Barcelona, Spain, 25–26 July 2004; pp. 404–411.

51. Nasar, Z.; Jaffry, S.W.; Malik, M.K. Textual keyword extraction and summarization: State-of-the-art. *Inf. Process. Manag.* **2019**, *56*, 102088. [CrossRef]

52. Campos, R.; Mangaravite, V.; Pasquali, A.; Jorge, A.M.; Nunes, C.; Jatowt, A. YAKE! Keyword extraction from single documents using multiple local features. *Inf. Sci.* **2020**, *509*, 257–289. [CrossRef]

53. Shouzhong, T.; Minlie, H. Mining microblog user interests based on TextRank with TF-IDF factor. *J. China Univ. Posts Telecommun.* **2016**, *23*, 40–46. [CrossRef]

54. Paramonov, I.; Lagutina, K.; Mamedov, E.; Lagutina, N. Thesaurus-Based Method of Increasing Text-via-Keyphrase Graph Connectivity During Keyphrase Extraction for e-Tourism Applications. In Proceedings of the Knowledge Engineering and Semantic Web; Ngonga Ngomo, A.-C., Křemen, P., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 129–141.

55. Gagliardi, I.; Artese, M.T. Semantic Unsupervised Automatic Keyphrases Extraction by Integrating Word Embedding with Clustering Methods. *Multimodal Technol. Interact.* **2020**, *4*, 30. [CrossRef]

56. Ali, F.; El-Sappagh, S.; Riazul Islam, S.M.; Kwak, D.; Ali, A.; Imran, M.; Kyung-Sup, K. A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion. *Inf. Fusion* **2020**, *63*, 208–222. [CrossRef]

57. Cheng, J.; Wang, Z.; Wen, J.; Yan, J.; Chen, Z. Contextual Text Understanding in Distributional Semantic Space. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, Melbourne, Australia, 19–23 October 2015; pp. 133–142.

58. Wang, J.; Wang, Z.; Zhang, D.; Yan, J. Combining Knowledge with Deep Convolutional Neural Networks for Short Text Classification. In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17), Melbourne, Australia, 19–25 August 2017; pp. 2915–2921.

59. Rosso, P.; Yang, D.; Cudremauroux, P. Revisiting Text and Knowledge Graph Joint Embeddings: The Amount of Shared Information Matters! In Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019.

60. Wu, W.; Li, H.; Wang, H.; Zhu, K.Q. Probase: A probabilistic taxonomy for text understanding. In Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data, Scottsdale, AZ, USA, 20–24 May 2012; pp. 481–492.

61. Microsoft Concept Graph and Concept Tagging Release. Available online: https://concept.research.microsoft.com/Home/Introduction (accessed on 16 June 2020).

62. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.S.; Dean, J. Distributed Representations of Words and Phrases and their Compositionality. In Proceedings of the Neural Information Processing Systems, Lake Tahoe, NV, USA, 5–10 December 2013; pp. 3111–3119.

63. Mikolov, T.; Chen, K.; Corrado, G.S.; Dean, J. Efficient Estimation of Word Representations in Vector Space. *arXiv* **2013**, arXiv:1301.3781.

64. Balli, S.; Karasoy, O. Development of content-based SMS classification application by using Word2Vec-based feature extraction. *IET Softw.* **2019**, *13*, 295–304. [CrossRef]

65. Dong, Y.; Liu, P.; Zhu, Z.; Wang, Q.; Zhang, Q. A Fusion Model-Based Label Embedding and Self-Interaction Attention for Text Classification. *IEEE Access* **2020**, *8*, 30548–30559. [CrossRef]

66. Liu, B. *Web Data Mining*; Springer Berlin Heidelberg: Berlin/Heidelberg, Germany, 2007; ISBN 978-3-540-37881-5.

67. Microsoft Knowledge Graph. Available online: https://concept.research.microsoft.com/Home/API (accessed on 17 June 2020).

68. SentiStrength. Available online: http://sentistrength.wlv.ac.uk/ (accessed on 16 June 2020).

69. Chawla, N.V.; Japkowicz, N.; Kotcz, A. Editorial: Special issue on learning from imbalanced data sets. *Sigkdd Explor.* **2004**, *6*, 1–6. [CrossRef]

70. Hu, Q.; Pedrycz, W.; Yu, D.; Lang, J. Selecting Discrete and Continuous Features Based on Neighborhood Decision Error Minimization. *IEEE Trans. Cybern.* **2010**, *40*, 137–150.

71. Lecca, M.; Rizzi, A.; Serapioni, R. GRASS: A Gradient-Based Random Sampling Scheme for Milano Retinex. *IEEE Trans. Image Process.* **2017**, *26*, 2767–2780. [CrossRef] [PubMed]

72. Manning, C.; Schutze, H. *Foundations of Statistical Natural Language Processing*; MIT Press: Cambridge, MA, USA, 1999.

73. Sindhu, I.; Daudpota, S.M.; Badar, K.; Bakhtyar, M.; Baber, J.; Nurunnabi, M. Aspect-Based Opinion Mining on Student's Feedback for Faculty Teaching Performance Evaluation. *IEEE Access* **2019**, *7*, 108729–108741. [CrossRef]

74. Dong, L.; Huang, S.; Wei, F.; Lapata, M.; Zhou, M.; Xu, K. Learning to Generate Product Reviews from Attributes. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, Valencia, Spain, 3–7 April 2017; Volume 1, pp. 623–632.