Proceedings of the 4th International Conference on Computing and Data Science (CONF-CDS 2022)
DOI: 10.54254/2755-2721/2/20220526

A combined analysis system for network products based on aspect extraction and sentiment classification

SONG Shangzheng

School of Electrical and Computer Engineering, Xiamen University Malaysia, 43900 Sepang, Negeri Selangor, Malaysia

cst1709333@xmu.edu.my

Abstract. With the continuous development of online shopping, more and more online shopping review data has been accumulated. These reviews contain consumers' attitudes and opinions on the products, which are of great value for mining. This research introduces a set of methods to mine review information by using aspect extraction and sentiment analysis. By analyzing the aspects involved in the reviews and the sentiment polarity of the reviewers, the potential opinions on the products can be summarized and presented to the merchants so as to understand consumers' preferences and make reasonable improvements to the products. This research combines and adopts several techniques, such as aspect extraction, paragraph vector and sentiment classification, to build a complete process which can handle review text and get the results of review aspects and sentiment polarity. By taking an Amazon review dataset as an example and applying this process, the relevant aspects and sentiment polarities contained in the reviews are obtained, which reveals that the idea of this research is feasible.

Keywords: sentiment analysis, aspect extraction, machine learning, paragraph vector.

1. Introduction

Nowadays, online shopping is on the rise and produces a huge number of products with customers' reviews. However, it is difficult to summarize information from such a large amount of non-numerical data, which leads to the waste of data to some extent.

This research attempts to combine a series of existing opinion mining and sentiment analysis methods to form a set of procedures to extract comment topics and sentiment polarities, so as to help producers better understand consumers' attitudes towards their products.

Although there are many papers that present advanced methods of aspect extraction or sentiment classification, most of them are dedicated to a single field. For example, some new methods are very effective in analyzing emotions, but there is no mention in the corresponding article of how to find the themes and topics corresponding to these emotions. The problem is that practitioners can easily find advanced techniques for natural language processing in papers, but it takes time to combine multiple techniques and consider the different effects of different techniques combined.

In this research, the author tries to combine different techniques, and summarizes an effective complete process from pre-processing to outputting the results of the relevant topics and sentiment polarity of the reviews, and uses sports game reviews obtained from Amazon shopping website for testing. Some of the work related to aspect extraction and sentiment classification will be reviewed in the section of

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2/20220526

Literature Review. Specific steps and methods in the whole process will be presented in the section of Methodology. The results of the models and related discussions will be placed in the Results and Discussions section. Finally, the summary of the study and the plan for future work will be provided in the Conclusion.

2. Literature review

In order to enable analysis of reviews, it needs to break the entire task down into several modules. A product can be judged from many different perspectives, so when analyzing a shopping review, it is important to know what the review is talking about. Therefore, first of all, an extractor is needed to extract the content mentioned in the comments. This process is usually called feature or aspect extraction. Comments that have gone through the aspect extraction step are labeled with the aspects of their content and then served as output for further processing. After a feature is obtained, a polarity analysis should be performed on the related comments to determine whether the customer is admiring the feature or expressing dissatisfaction with that aspect. Therefore, a sentiment classifier is needed to analyze the comments and output whether each of them is positive or negative. Based on the above operations, feature/aspects extraction and sentiment classification will be the two main technologies of this project and discussed in the following.

2.1. Aspect extraction

Aspect extraction is an important step in opinion mining. Opinion mining usually refers to the extraction of the author's ideas or attitudes towards the topic of the text from some unstructured texts, especially subjective texts such as reviews. While feature extraction is to identify the topic or characteristics involved in the text first, as the basis for the subsequent analysis of emotions or attitudes.

For example, the sentence 'It makes me feel excited and intense, because its storyline is coherent and tight' includes the aspect 'storyline' of a game. If the machine can automatically recognize this feature with natural language processing technology, it is called aspect extraction.

In the early days, for some text content without domain knowledge, the method of calculating the similarity between texts can be used to find similar content. Hofmann [9] developed a framework to seek the similarities between documents with some statistical methods which support unlabeled text.

Later came a method for classifying text topics called the topic model. McDonald and Titov [14] developed a multi-grain topic model which can not only identify important terms, but also cluster them into coherent groups based on some concept of Latent Dirichlet Allocation (LDA) [3] and Probabilistic Latent Semantic Analysis (PLSA). The model extracted the multi-grain outputs and evaluated it quantitatively and qualitatively to show the performance improvement compared with the standard models. Zhao et al. [18] proposed a MaxEnt-LDA hybrid model to find both aspects and aspects-specific opinion words jointly, and this topic model can effectively identify aspects and opinion words simultaneously on a relatively small amount of training text data. Zou et al. [19] also used topic model to extract product features and appended the process of TFIDF-filter and a novel product similarity calculation that performs weighted fusion based on information entropy, and finally enhanced the performance of the extraction effectively.

The traditional topic model related to LDA is influenced by the length of the article, and it is often unable to achieve good performance in short texts. Therefore, some topic models specifically for short texts have emerged. Yan et al. [17] proposed Bi-term topic model (BTM) which learns the topics by directly modeling the generation of word co-occurrence patterns in the whole corpus. The results showed that this approach can find more prominent and coherent topics, and significantly outperform baseline methods on several evaluation metrics. In addition to BTM, a new approach named Sentence Segment LDA (SS-LDA) was proposed by Ozyurt and Akcayol [13], which improved the process of segmentation of sentence and was competitive in extracting product features.

DOI: 10.54254/2755-2721/2/20220526

2.2. Sentiment classification

The construction of sentiment classifier is another important step in opinion mining, which is mainly to choose the appropriate method to judge the sentiment tendency of the obtained text. In the early stage, people acquired the emotional score of each sentence or article by constructing emotion dictionaries. With the development of machine learning theory, different traditional machine learning algorithms have been applied to sentiment analysis. However, in recent years, more and more sentiment classifiers have begun to use deep learning algorithms based on neural networks. Each of these approaches has its own advantages and disadvantages.

2.2.1. Methods based on sentiment dictionary. Sentiment analysis using a sentiment dictionary [5] refers to extracting sentiment words from sentences or articles, and calculating the score of these sentiment words in sentiment dictionaries according to certain rules, so as to get the sentiment score of the whole sentence or article. The advantage of this method is that it does not need a training set with data annotation, but the difficulty lies in how to assign appropriate scores and weights to each emotion word.

2.2.2. Methods based on traditional machine learning. Due to the particularity of language features, people generally use some classic algorithms, such as Support Vector Machine (SVM), Naïve Bayes and Logistic Regression, to carry out traditional machine learning on sentiment classification.

In order to improve the computational efficiency and scaling ability of large-scale reviews, Xu et al. [16] proposed a continuous naive Bayes learning framework for sentiment classification of product reviews on large-scale multi-domain e-commerce platforms. This framework extended the parameter estimation mechanism of Naive Bayes model to a continuous learning method while maintaining the high computational efficiency of traditional Naive Bayes model. The experimental results on Amazon product review datasets showed that the model is better able to process constantly updated reviews from different fields.

To process a large volume of text data, including more than 4 million product reviews from Amazon, Al-Saqqa et al. [1] used three learning methods, SVM, NB and LR, to train the model in a scalable data-processing system, Apache Spark. Their result showed that SVM has better accuracy (86%) than the other two methods (85.4% and 81.4%) under the current conditions.

Tripathy et al. [15] used more traditional machine learning methods to train the IMDB dataset, and tried to arrange and combine different values of n in n-gram with these learning methods, and then evaluated them from multiple perspectives such as precision, recall, F-measure and accuracy, finally got a more detailed result. Generally, using the method of 'Unigram + Bigram' and SVM achieved the best results, with an accuracy rate of 88.94 %.

The traditional word representations are bag-of-words (BOW) and term frequency-inverse document frequency (TFIDF), but their defects lie in that they do not involve word order, context and semantic relations and other information. Word2vec [11] based on continuous bag-of-words (CBOW) or skip-gram is a tool that can capture context and semantic similarities in data. Bansal and Srivastava [2] used word2vec for preprocessing, and used Naive Bayes, SVM, Logistic Regression and Random Forest to train the review data set obtained from Amazon, and constructed the sentiment classifier. Their results showed that CBOW performed better on this dataset than skip-gram. On this basis, the classifier accuracy trained by Random Forest is the highest, reaching 90.66 %. SVM and LR have a slightly lower score, while NB lags far behind at about 55 %.

2.2.3. Methods based on deep learning. With the continuous development of deep neural network theory, more and more people begin to combine sentiment analysis and deep learning for research. Although traditional machine learning methods have been able to achieve high accuracy, sometimes some reasonable deep neural network architectures could get better results.

There are two architectures that are currently more commonly used for natural language processing, that is, Convolutional Neural Networks (CNN) and long-term short-term memory recurrent Neural Net-

DOI: 10.54254/2755-2721/2/20220526

works (LSTM). Colon-Ruiz and Segura-Bedmar [6] combined CNN and LSTM with different preprocessed word embedding models, and the hybrid model composed of a bidirectional LSTM followed by a CNN is found to be the best. In addition, they tried the Bidirectional Encoder Representations from Transformers (BERT) with a Bi-LSTM to analyze the sentiment analysis of drug reviews and found that the Bert model worked very well, but training took a long time. In comparison, CNN's results were slightly worse but acceptable, and required less training time.

3. Methodology

During the design and implementation of the entire study, an actual data set was needed to test the feasibility of the process and select highly accurate training methods, so we used product reviews from Amazon's sports video game category as an example to conduct the entire process. However, the raw data still needs to be processed by some techniques of natural language processing to obtain appropriate results that can be analyzed. The pre-processed data will be used to extract aspects and judge sentiments, which will be the focus of this section. After obtaining the results, a visualization step could be optionally added for users to analyze and find some trends. Therefore, the entire workflow of processing and analysis can be divided into six main modules, namely data collection and formatting, data preprocessing, aspect extraction and allocation, document vector modeling, sentiment classification, result visualization and trend analysis.

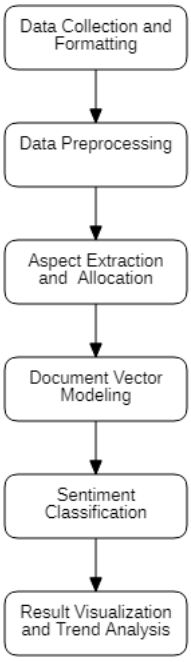

Figure 1. Overall workflow of the process.

As shown in Figure 1, the data is collected and formatted into a structure that will then be converted into a form acceptable to models using some pre-processing techniques. Next, a topic model will be constructed to extract the potential topics and after the extraction, and this model can allocate the text to which aspects it belongs to. The following step is building a Doc2vec model. In this module, the original documents will be converted into document vectors and word vectors. This distributed representation replaces the simple bag of words (BOW) and can be used as input to the sentiment classifier in the sentiment classification section. Later, a variety of machine learning algorithms will be tried to construct the classifier and the most appropriate one will be selected as the final algorithm for sentiment classification. Finally, the results of aspect extraction and sentiment classification can be visualized by existing tools and the trends will be analyzed based on the generated visual charts.

DOI: 10.54254/2755-2721/2/20220526

3.1. Data collection and formatting

The data used in this study came from three different sources for different purposes.

One piece of data is used to train the classifier. In order to adjust the model parameters of the supervised learning algorithm, the data needs to be manually labeled correctly, so it needs to have not only the review text, but also the corresponding tagged sentiment labels, and it had better be contextually relevant. The Multi-Domain Sentiment Dataset [4] we used in this study meets the above requirements.

The second data is word corpus data, which is used to expand the vocabulary of the model during training to reduce the possibility of failure to identify words when predicting new documents. We chose the data from the Amazon Review Data [12] as the word corpus data because it provides many different categories of Amazon product reviews, including the corpus of sports video game reviews that meet the needs of this study.

Another piece of data is the data to be analyzed. This piece of data will be input into the trained model and output results, which will be used as the data basis for analysis. Reviews of sports video games on Amazon will be tested as an example in this study, and the following work will be based on the above three kinds of data.

It should be noted that when the data to be analyzed changes, the data for training classifier and the word corpus data should also be adjusted or replaced in accordance with the actual situation.

3.2. Data preprocessing

In this research, there are different preprocessing techniques for different data.

For the data used to train the classifier and to serve as a corpus, only a few simple processing rules are required, such as removing short text, converting punctuation and special characters, and converting the case.

The review text to be analyzed will be input to the Aspect Extraction module for classification, but the module requires a sparse matrix with textual characteristic values as input. In order to convert the document from textual format to numeric format, a further pre-processing process is required, including text tokenization, removal of stop words, generation of vocabulary and generation of document-term matrix.

3.3. Aspect extraction and allocation

Before sentiment classification, aspect extraction is required. ‘Aspect’ here refers to the main directions or angles covered in a document. For the reviews to be analyzed, in order to categorize them according to the text content, it is necessary to know the aspects contained in the text before performing the following analysis.

DOI: 10.54254/2755-2721/2/20220526

Due to the lack of specific category-labeled data that can be used to train the model, a supervised algorithm cannot be used to classify the review text. Topic model is a statistical model that can cluster the latent semantic structure of corpus based on unsupervised or semi-supervised algorithms. In this study, a semi-supervised topic model based on Correlation Explanation (CorEx Topic Model [8]) is used to extract the aspects in reviews. The related Work Scenario is shown in Figure 2.

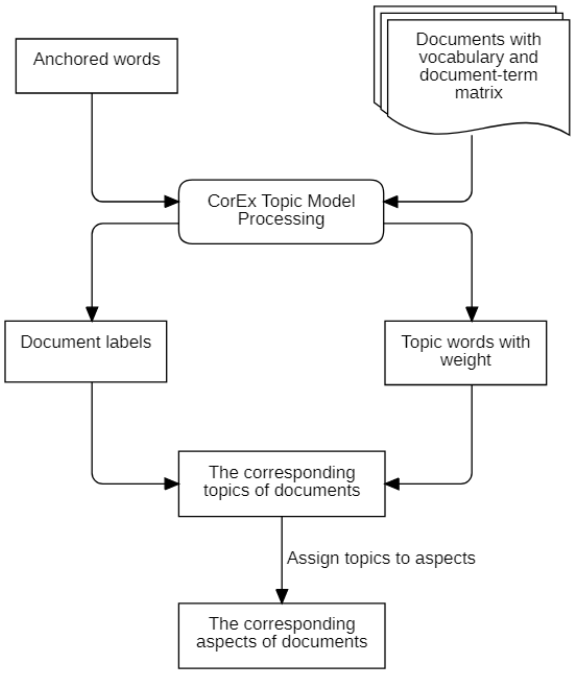

Figure 2. The work scenario of aspect extraction and allocation.

As can be seen from the scenario, documents with vocabulary and document-term matrix and the anchored words (which will be explained in the section of Explanation of Anchored word) are entered into the model. Documents with vocabulary and document-term matrix is the formatted data previously obtained in the data preprocessing module, while anchored words are summed up based on priori domain knowledge.

After running the model, two pieces of data will be generated. One piece of data is called document labels, which are used to record which topic each document belongs to, and is represented by True or False. The other data is called topic words with weight, which records the words and their weights in each topic. These two pieces of data are combined and artificially assign aspects to topics by looking at the words and weights in each topic to get the aspect to which each review/document belongs. In actual shopping reviews, consumers may comment from multiple perspectives, so a review may belong to multiple topics, which also means that the review may belong to multiple aspects.

3.3.1. Explanation of anchored word. A semi-supervised topic model can be implemented by anchored word. The user can specify the topic words that are generated based on prior knowledge so that the vocabulary generated by the model is related to the topic related to the anchored word.

In other words, the explanatory topic that the user wants could be generated by adding anchored words for the specified topic.

3.4. Document vector modeling

Although one-hot notation is the most common and fundamental method used to process textual information, it still has many drawbacks. For example, it is a sparse representation that could lead to a dimensional disaster, and each word represented by this method is independent and fails to reflect the

DOI: 10.54254/2755-2721/2/20220526

implied semantic relationship. Therefore, in this research, the doc2vec method is adopted to get the vector of documents and words for the purpose of the following sentiment classification.

Doc2vec, also known as paragraph vector or sentence embedding, is a distributed representation method of documents and sentences and it was introduced by Mikolov and Le [10] on the basis of the famous word embedding method, word2vec. It adds a paragraph vector to word2vec, which is used to represent the features of the paragraph content. The workflow diagram of the module in this research is shown in Figure 3.

As can be seen from Figure 3, three pieces of data are input into the tag function. These three data have been formatted pre-processing. Among them, reviews are the text to be analyzed that has gone through the Aspect Extraction and Allocation module, and training set is the labeled review text obtained from the Multi-Domain Sentiment Dataset, which its resulting vectors will be used to train the classifier. The Corpus is a video game-related review dataset obtained from Amazon, which is used to extend more video game-related vocabulary to the model.

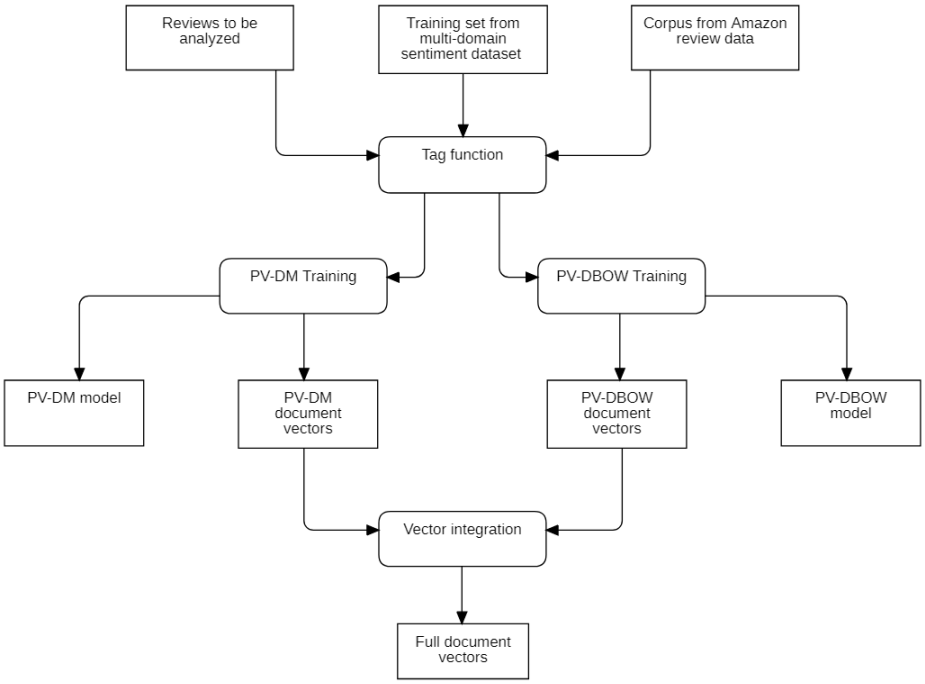

Figure 3. The process of getting doc2vec model and obtaining document vectors from datasets.

When the data is entered into the tag function, the function breaks each paragraph into words according to the whitespace and assigns a unique tag. This tag acts as an identifier. After the model is trained, it will be used to identify the document represented, so that the trained document vector can be easily obtained from the model. After tagging, data will be input into the PV-DM (Distributed Memory Model of Paragraph Vectors [10]) model and PV-DBOW (Distributed Bag of Words of Paragraph Vectors [10]) model respectively for training. After the training is completed, the two processes will generate their own model and document vectors respectively. The final complete document vectors can be obtained by combining the two pieces of vectors together through vector integration. This full document vectors are the paragraph vector representation of the document transformed.

DOI: 10.54254/2755-2721/2/20220526

In addition, the model training process produces two models with internal parameters that can be used to transform the new unknown document into paragraph vectors. The specific process is shown in Figure 4.

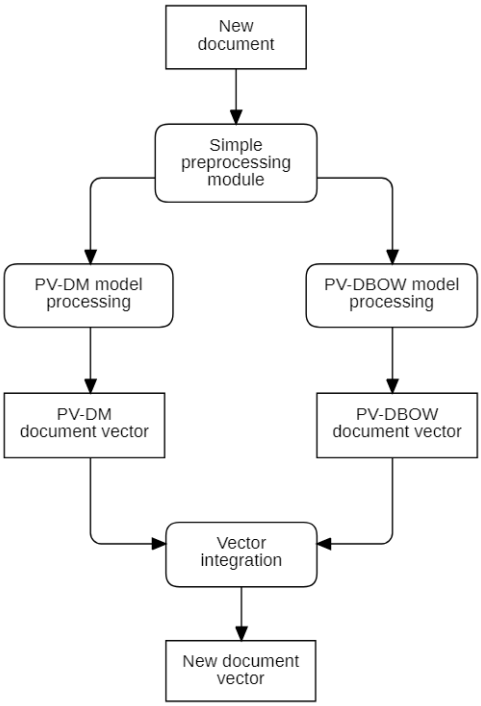

Figure 4. The process of new document paragraph vector generation.

In this process, the document will be processed by the simple preprocessing module first, and then two vectors will be generated by two models respectively, and the final paragraph vector results will be generated through vector integration without training again.

3.5. Sentiment classification

After obtaining the paragraph vectors represented by numerical values, different algorithms can be easily applied to them to train the sentiment classifier. In this study, three different traditional machine learning algorithms (Naive Bayes, Logistic Regression and Support Vector Machine) will be evaluated for sentiment classification of reviews.

4. Results and discussions

In the actual implementation, a total of two models and a classifier (containing three algorithms) were established, of which the CorEx topic model and classifier were used to output the results of the review text, so they will be evaluated in the following section. However, the Doc2vec model is not used to output the results of aspect or sentiment, but as a means to convert text into vector, so it is not within the evaluation scope of this research.

4.1. Evaluation of CorEx Topic Model

In order to evaluate CorEx Topic Model, one of the model authors (Ryan J. Gallagher) mentioned in the GitHub [7] about this project that they usually use Total Correlation (TC) between models to evaluate an existing model, rather than an absolute score. Topic with higher TC will 'explain' more about the collection of documents. If the overall TC does not increase significantly after the addition of a topic, then the existing topic may already contain most related words that can explain the document.

DOI: 10.54254/2755-2721/2/20220526

The data to be analyzed in this study is a collection of 8491 user reviews from sports video games on Amazon, and there were sixteen groups of anchored words used, so these reviews will form at least 16 topics. In order to verify whether these topics can cover the review documents, we increased the number of topics of the model one by one, from 16 to 24, and calculated corresponding TC values. The overall TC of each version of model is shown in Table 1.

Table 1. The test result of topic numbers from 16 to 24.

The number of topics	Total Correlation
16	30.43
17	30.50
18	31.04
19	30.90
20	31.10
21	31.06
22	30.66
23	31.27
24	31.51

As can be seen from Table 1, with the increase in the number of topics, overall TC did not increase significantly (from 30.43 to 31.51), and even decreased in stages 18-19 (from 31.04 to 30.90) and 20-22 (from 31.10 to 30.66). This indicates that when the number of topics is 16, the existing topic words can explain the word relations in the documents well. However, the final number of topics was chosen at 18, as a precaution against the very small possibility of potential topic words not being included in the model. The training time for the model with 16 topics was 44.23 seconds, compared to 56.96 seconds for the model with 18 topics. The 12.73 seconds is not a big burden for the whole research process, so the time cost caused by adding two additional topics is acceptable. Finally, the 18 groups of words obtained from the model were grouped into five general aspects, namely, ‘Game Quality’, ‘Shopping Experience’, ‘Game Content’, ‘Device & Hardware’ and ‘Others’.

However, comparing TC values can only select a reasonable number of topics, but not reflect the performance of the model. Thus, 104 reviews containing one aspect were randomly selected and then artificially judged which aspect belonged to. By comparing the predicted aspects of the model with the manually annotated aspects, a classification report of the model can be generated. The aspect extraction results for each aspect are given in Fig. 5.

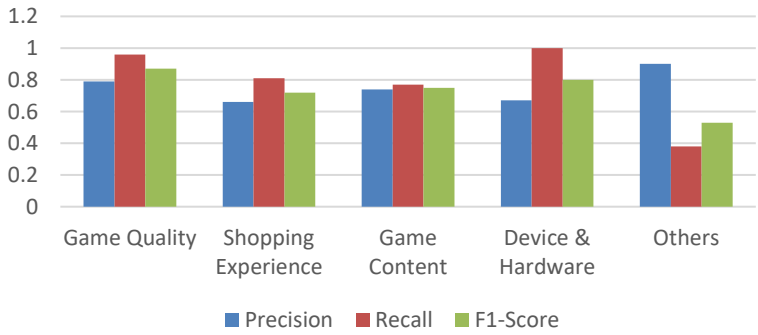

Figure 5. Evaluation result of CorEx topic model.

As can be seen from Figure 5, the model has good precision, recall and f1-score in aspects of ‘Game Quality’, ‘Shopping Experience’, ‘Game Content’ and ‘Device & Hardware’. However, in the aspect of ‘Others’, although the precision is very high, the recall rate is obviously low, which means that many reviews actually belonging to the aspect of ‘Others’ are identified as other aspects. We think the reason

DOI: 10.54254/2755-2721/2/20220526

for this may be that the 'Others' aspect is a complementary aspect intended to include all the reviews that aren't part of the other four aspects, but when using the anchored word, these words are all classified as the first four aspects, and there is no anchored word specially belonging to 'Others' aspect. Therefore, when the reviews actually belong to the aspect of 'Others', the model tends to identify them as these four aspects, resulting in a low recall rate in 'Others' aspect.

Also, it should be noted that due to limited time and resources, only 104 pieces of data were manually annotated as the test set in this research. Therefore, the resulting report has certain limitations and it is only suitable for the data reference in this implementation.

4.2. Evaluation of sentiment classifier

A total of 38 548 reviews with positive or negative tags were obtained from the Multi-Domain Sentiment Data Set. The training time and evaluation results for classifiers of different algorithms are shown in Table 2.

Table 2. Training time for classifiers based on different algorithms.

Algorithm	Computing Time	Accuracy
Naïve Bayes	0.43 Seconds	0.79
Logistic Regression	2.40 Seconds	0.85
Support Vector Machine	24.5 Minutes	0.87

Combined with computing time and accuracy of models, the most appropriate algorithm in different scenarios can be selected flexibly. For example, for the review data in this study, the Naïve Bayes algorithm will not be considered, because although its running speed is fast, its accuracy is obviously lower. For the other two algorithms, Logistic Regression can be selected when testing the model and the feasibility of the process, because of its high accuracy and significantly shorter computing time than SVM. When the process is complete and the model is applied to judge sentiments, SVM is the wiser choice because it has the highest accuracy and the training time of the SVM is only about 25 minutes, which is completely acceptable.

4.3. Discussion of selection of aspect extraction model

In the implementation process, the biggest issue encountered is that the topic model cannot effectively gather the topic words. At present, topic model is a popular method to extract the aspects involved in paragraphs. However, in the initial experiment, due to the limitations of unsupervised algorithm, the topic words generated by this model could not form obvious topics. Even when the stop list filtered out the nonsense words, the results were still not good.

It took a long time to realize the aspect extraction with good performance. During this period, topic models based on different algorithms (Latent Dirichlet Allocation, Non-negative Matrix Factorization and Correlation Explanation) were repeatedly tested, and it was finally found that if the prior knowledge was added to the model based on semi-supervised algorithm, it could effectively guide the model to generate different meaningful topics and achieve a good effect. Therefore, the CorEx topic model with anchored words was eventually used in the framework.

4.4. Discussion of selection of dataset for classifier training

In the Amazon Review Dataset, there is a column of data called 'Rating', which records the star rating of each review and is used to show consumer satisfaction with a product. It ranges from '1' to '5' and the higher the star rating, the more satisfied the consumer is. Therefore, in the initial idea, Amazon Review Dataset could be used as the Data set of the classifier training. The 'Rating' of each review in the data set would be used as the label of the sentiment tendency of the review, that is, the reviews with a star rating of '5' would be regarded as positive reviews, and reviews with a star rating of '1' as negative reviews. This strategy seems reasonable.

DOI: 10.54254/2755-2721/2/20220526

At the very beginning, when the training of the classifier by the Amazon Review Dataset was completed, the results it presented were exciting. The highest accuracy rate reached 94 % when using SVM algorithm, which is quite a high number. However, in the following validation, there were some problems with this classifier. When predicting sentences that are easy to detect sentiment tendencies, such as 'I don't like this stuff very much', the model gave wrong results. After thinking and verifying from multiple perspectives, the conclusion was that the problem should appear in the selected data set. In Amazon Review Data, 'Rating' could indeed reflect consumers' sentiment and opinions to a certain extent. However, as a text with variable length and completely composed of personal opinions, reviews often contain more than one kind of emotional vocabulary. For example, even in a five-star review, after evaluating the advantages of the product, the reviewer might still point out the shortcomings of this product. Therefore, in the model training, these negative fragments would be regarded as positive sentiment, which may lead to fallacies.

Under this circumstance, the dataset of the training classifier is replaced by the Multi-Domain Sentiment Dataset. The sentiment labels of this data set were all manually reviewed and labeled, so the proportion of reviews containing double sentiments was very low. Although the accuracy of the classifier decreased after the completion of training (0.87), it almost did not misjudge sentences with obvious sentiment tendencies during verification.

5. Conclusion

In this research, an experiment was achieved to process video game review content through aspect extractor and sentiment classifier, and to analyze the hidden aspects and emotions in reviews. Although the reviews used in this experiment are limited to the field of sports video games, in fact, the models and methods used in the whole process form a workflow for handling reviews, which can be reused to reviews from other fields. In this way, sellers in a particular domain can use the framework to conduct analysis for that domain by only changing the training dataset and the priori knowledge (anchored words). Through certain analysis, sellers can have a clearer and in-depth understanding of the current trends in the field, so as to purposefully improve the characteristics of their products and increase the competitiveness of their products.

We provide a practical example here: https://public.tableau.com/app/profile/ssz1697/viz/2_16254581776420/1. All the data used in this Tableau-generated-example is from the data set or model results used in this study. It shows statistics and trends about aspects and sentiments contained in the reviews from different perspectives, which shows that this workflow can be expanded to provide more possibilities for users to analyze reviews of different kinds of products.

There are some limitations in this research, especially the part about topic model. In order to generate a well-performing topic model, the anchored words should be input to the model. Naturally, for single-domain reviews, we can obtain appropriate anchored words based on prior knowledge. However, if the reviews cover a wide range of areas, the normal number of anchored words may not fully cover the topics, resulting in performance degradation of the model. Therefore, in the future, some other unsupervised topic models could be tried, and if there is another model that achieves the same accuracy and does not require anchored words, it will obviously reduce the difficulty of aspect extraction.

References

[1] Al-Saqqa, S., Al-Naymat, G., & Awajan, A. (2018). A large-scale sentiment data classification for online reviews under apache spark. Procedia Computer Science, 141, 183-189.

[2] Bansal, B., & Srivastava, S. (2018). Sentiment classification of online consumer reviews using word vector representations. Procedia computer science, 132, 1147-1153.

[3] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. the Journal of machine Learning research, 3, 993-1022.

[4] Blitzer, J., Dredze, M., & Pereira, F. (2007) 'Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification' Proceedings of the 45th annual meeting of the association of computational linguistics. pp. 440-447.

DOI: 10.54254/2755-2721/2/20220526

[5] Chen, C.-C., Huang, H.-H., & Chen, H.-H. (2018) 'NTUSD-Fin: a market sentiment dictionary for financial social media data applications' Proceedings of the 1st Financial Narrative Processing Workshop (FNP 2018).

[6] Colon-Ruiz, C., & Segura-Bedmar, I. (2020). Comparing deep learning architectures for sentiment analysis on drug reviews. J Biomed Inform, 110, 103539. doi:10.1016/j.jbi.2020.103539.

[7] Gallagher, R. J. (2021). Metrics for Model Selection. https://github.com/gregversteeg/corex_topic/issues/51. Accessed 28 July, 2021.

[8] Gallagher, R. J., Reing, K., Kale, D., & Ver Steeg, G. (2017). Anchored correlation explanation: Topic modeling with minimal domain knowledge. Transactions of the Association for Computational Linguistics, 5, 529-542.

[9] Hofmann, T. (2000) 'Learning the similarity of documents: An information-geometric approach to document retrieval and categorization' Advances in neural information processing systems. pp. 914-920.

[10] Le, Q., & Mikolov, T. (2014) 'Distributed representations of sentences and documents' International conference on machine learning. PMLR, pp. 1188-1196.

[11] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013) 'Distributed representations of words and phrases and their compositionality' Advances in neural information processing systems. pp. 3111-3119.

[12] Ni, J., Li, J., & McAuley, J. (2019) 'Justifying recommendations using distantly-labeled reviews and fine-grained aspects' Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). pp. 188-197.

[13] Ozyurt, B., & Akcayol, M. A. (2021). A new topic modeling based approach for aspect extraction in aspect based sentiment analysis: SS-LDA. Expert Systems with Applications, 168. doi:10.1016/j.eswa.2020.114231.

[14] Titov, I., & McDonald, R. (2008). Modeling online reviews with multi-grain topic models. (Paper presented at the Proceedings of the 17th international conference on World Wide Web, Beijing, China).

[15] Tripathy, A., Agrawal, A., & Rath, S. K. (2016). Classification of sentiment reviews using n-gram machine learning approach. Expert Systems with Applications, 57, 117-126. doi:10.1016/j.eswa.2016.03.028.

[16] Xu, F., Pan, Z., & Xia, R. (2020). E-commerce product review sentiment classification based on a naïve Bayes continuous learning framework. Information Processing & Management, 57(5). doi:10.1016/j.ipm.2020.102221.

[17] Yan, X., Guo, J., Lan, Y., & Cheng, X. (2013). A biterm topic model for short texts. (Paper presented at the Proceedings of the 22nd international conference on World Wide Web, Rio de Janeiro, Brazil).

[18] Zhao, X., Jiang, J., Yan, H., & Li, X. (2010) 'Jointly modeling aspects and opinions with a MaxEnt-LDA hybrid'. ACL.

[19] Zou, X., Zhu, Y., Feng, J., Lu, J., & Li, X. (2019). A novel hierarchical topic model for horizontal topic expansion with observed label information. IEEE Access, 7, 184242-184253.