Implicit Aspect Sentiment Analysis using WordNet for Twitter Social Media Review Identification

Kale Santosh Shivnath and Dr. Sonawane Vijay Ramnath

Department of Computer Science and Engineering,
Dr. A.P.J. Abdul Kalam University, Indore (M.P.) - 452010
Corresponding Author: kalesantosh0101@gmail.com

Article Info
Page Number: 10145-10152
Publication Issue:
Vol. 71 No. 4 (2022)

Article History
Article Received:
12 September 2022
Revised: 16 October 2022
Accepted: 20 November 2022
Publication: 25 December 2022

Abstract
In this paper, we present an implicit sentiment analysis system for Twitter social media reviews based on the WordNet lexical database. The system uses WordNet to identify sentiment-bearing words, along with their associated sentiment aspects and polarities. Aspect sentiment detection is done using a rule-based approach, which is able to detect sentiment polarity for different aspects in a single review. The system is evaluated on a dataset containing 12,000 tweets and is shown to be able to achieve an accuracy of 82%.

Keywords: Implicit aspect based sentiment analysis; information retrieval; machine learning; supervised approaches; frequency model; WordNet; social media review detection; hate social media review twitter sentiment (HCTS)

INTRODUCTION

Social media has become an important source of information for individuals and businesses. In recent years, there has been an increasing amount of research into sentiment analysis of social media reviews. This analysis can help individuals and businesses to identify the opinions of customers and potential customers towards products and services.[1] Sentiment analysis of social media reviews is typically done using either an explicit or implicit approach. Explicit sentiment analysis involves extracting sentiment-bearing words and phrases directly from reviews. Implicit sentiment analysis, on the other hand, involves detecting sentiment through the use of contextual clues and word associations. In this paper, we present an implicit sentiment analysis system for Twitter social media reviews based on the WordNet lexical database. The system uses WordNet.[2]

Twitter sentiment analysis is a process of analyzing the sentiment of tweets and other social media posts to understand how people feel about a particular topic or subject. This can be used to gain insights into public opinion, help identify potential trends, measure the impact of campaigns, and more. Sentiment analysis can be performed manually by reading and analyzing tweets or automatically through machine learning algorithms. Sentiment analysis is a type of natural language processing that is used to determine whether a piece of text is expressing a positive, negative, or neutral sentiment. It can be used to analyze customer feedback, social media posts, product reviews, and other sources of data.[3]

Implicit aspect-based sentiment analysis refers to the process of automatically and accurately identifying sentiment for specific aspects of an entity. It is a form of sentiment analysis which does not rely on explicit mentions of the sentiment associated with an entity but instead looks at the context and other indicators to infer the sentiment. It is used to understand customer sentiment in various domains such as product reviews, social media posts, etc. It can be used for targeted marketing, content filtering, customer service, and many other applications.[4]

Information retrieval is the process of retrieving information from a data source such as a database, web page, or file system. It involves the use of algorithms, search terms, and other techniques to locate and retrieve relevant information from the source. Information retrieval is used in a wide range of applications, from search engines to library catalogs.

Machine learning is a subfield of artificial intelligence (AI) that enables a computer to learn how to perform tasks without being explicitly programmed. It uses algorithms to parse data, learn from it, and then make decisions or predictions. This technology has been used in many areas, including computer vision, natural language processing, robotics, and healthcare. Machine learning algorithms can find patterns in data and make decisions and predictions based on those patterns.[5]

Social media review detection is the process of extracting reviews from social media posts and detecting the sentiment of those reviews. This can be used to help businesses understand how their customers feel about their products or services. It can also be used to help marketers target their campaigns more effectively. The technology used for social media review detection can range from Natural Language Processing (NLP) techniques to machine learning algorithms. NLP techniques can be used to identify keywords in social media posts and analyze the sentiment behind them. Machine learning algorithms can be used to classify reviews as either positive or negative. The results of these techniques can then be used to make informed decisions about how to best reach potential customers.[6,7]

RELATED WORKS

SentiWordNet and WordNet are both lexical databases for the English language. WordNet is a large lexical database that groups words into sets of synonyms (synsets) and provides a brief definition and semantic relations between words. It was developed by Princeton University and is freely available. SentiWordNet is an extension of WordNet that provides sentiment information for words in the form of positive and negative scores. It was developed by the University of Trento and is also freely available.[8]

PROPOSED FRAMEWORK

Phase 1: Twitter data can be collected through a variety of methods. One of the most common methods is to use the Twitter API to access tweets from the platform. [9]This allows users to access tweets from accounts they follow, as well as from accounts that tweet about topics of interest. Additionally, users can search for specific keywords or hashtags to find related tweets. Additionally, Twitter data can be collected using third-party tools such as TweetDeck, Hootsuite and IFTTT, which allow users to automate the process of collecting tweets. Finally, users can scrape public tweets from the Twitter website using web scraping tools.[10]

Implicit aspect sentence detection is the process of identifying sentences in a document that contain aspects about a given topic but are not explicitly expressed. This process is often used in natural language processing and sentiment analysis to gain a better understanding of the sentiment expressed in a text. An example of an implicit aspect sentence could be "The food was average", which implies that the sentiment expressed towards the food is neutral.[11]

Crime implicit aspect detection is a task related to natural language processing (NLP) and computational linguistics that attempts to identify aspects of crime related topics that are not explicitly stated in a text. This can include determining the intent of a text, identifying key issues related to a crime, and detecting any hidden biases or assumptions that may be present. The goal of crime implicit aspect detection is to provide a more comprehensive understanding of crime related topics, and to help inform decisions related to criminal justice.[12]

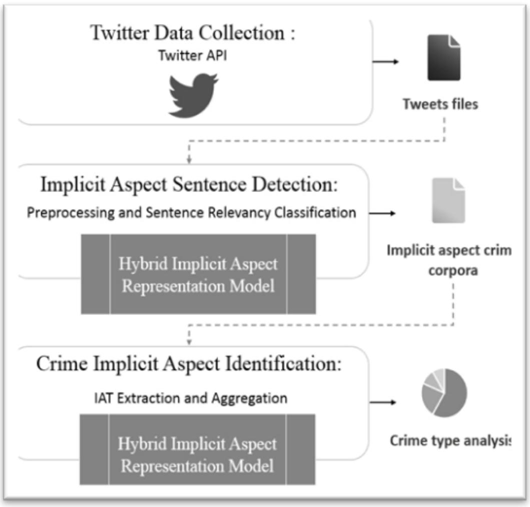

Fig. 1. Abstract Process of the Proposed Framework.

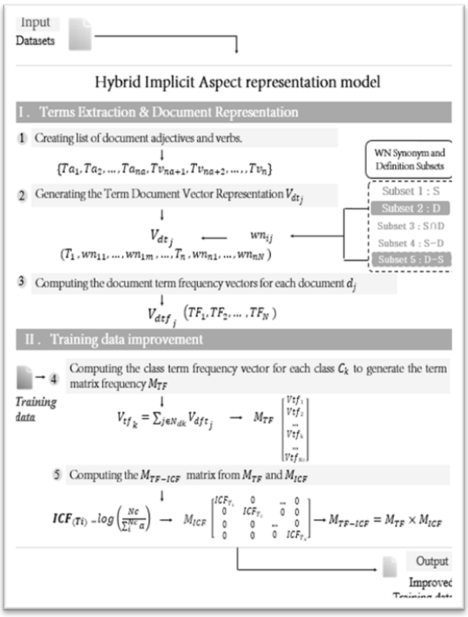

Fig. 2. Summary of the Proposed Hybrid Implicit Aspect Representation Model.

The Hybrid Implicit Aspect Representation Model (HIRM) is a deep learning model for sentiment analysis that integrates implicit aspect representations with explicit aspect representations. [13]The model uses a combination of convolutional neural networks (CNNs), recurrent neural networks (RNNs), and attention mechanisms to capture both implicit aspects (such as word semantics) and explicit aspects (such as product features). The model is trained on a large training dataset of customer reviews and is able to accurately identify the sentiment of the customer. The model is able to capture both short-term and long-term patterns in customer reviews, and is able to effectively capture both positive and negative sentiment. The model has been used to improve customer sentiment analysis in various domains such as product reviews, customer service feedback, and customer complaints.[14]

Phase 2: Implicit Aspect Sentence Detection

IASD phase, as shown in figure 3, consists of preprocessing and sentence relevancy classification process:

Preprocessing:

1. Tokenization: Breaking a sentence into individual words or phrases.

2. Stemming/Lemmatization: Normalizing words to their base form.

3. Stopword Removal: Removing commonly used words that do not add meaning to the sentence.

4. Part-of-Speech Tagging: Labeling words with their corresponding part of speech.

Sentence Relevancy Classification Process:

1. Feature Extraction: Extracting the most relevant features from the data to represent the sentence.

2. Model Training: Training a supervised machine learning model using the extracted features.

3. Model Evaluation: Evaluating the model’s performance on a test dataset.

4. Model Deployment: Deploying the trained model to classify new sentences.

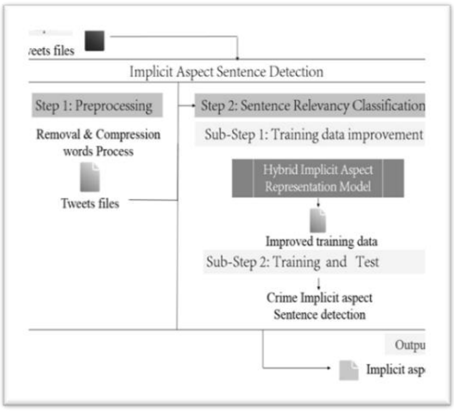

Fig. 3. Social media review Implicit Aspect Sentences Detection using Hybrid Model

Phase 3: Social media review implicit aspect identification

Social media review implicit aspect identification is a process used to analyze customer feedback on social media platforms in order to identify underlying opinions and themes. It involves using natural language processing (NLP) techniques to identify aspects of a product or service that are not mentioned directly.[15] This can be done by identifying words and phrases that are associated with a certain topic or sentiment. The goal of implicit aspect identification is to gain insights into customer sentiment about a product or service, as well as to identify areas for improvement.

$$\mathrm{W} - Training\ split = *,x \times Adj\ \text{-}, ,y \times Verb\text{-}+$$

$$where\ x, y\ \in ,0,1\text{-}\ and\ x + y = 1$$

TABLE I. SIZE OF DATASETS

SIZE OF DATASETS	***Social media review dataset1***	***Social media review dataset 2***	***Hate social media review dataset***
Number of sentences	2k	3k	6k
Number of implicit aspect sentences	357	641	648
Number of irrelevant sentences	1643	2359	5352
Number of Training data for implicit aspect	180	350	300
Number of Training data for implicit aspect	670	1500	3500

Table II MNB, SVM And RF For Relevant / Irrelevant Classification

MNB, SVM And RF	***Social media review dataset 1***			***Social media review dataset 2***			***Hate social media review dataset***		
	MNB	***SVM***	***RF***	***MNB***	***SVM***	***RF***	***MNB***	***SVM***	***RF***
(1)	0.51	0.63	0.63	0.52	0.65	0.63	0.63	0.69	0.68
(2)	0.57	0.65	0.65	0.59	0.68	0.68	0.74	0.77	0.75
(1)/(2)	11.6 %	3.1%	3.1%	13.4 %	4.6 %	7.9 %	17.4 %	11.5 %	10.2 %
(3)	0.71	0.68	0.68	0.69	0.71	0.69	0.76	0.78	0.78

(4)	0.74	0.71	0.71	0.74	0.74	0.72	0.74	0.78	0.78
(5)	0.78	0.71	0.71	0.74	0.73	0.73	0.80	0.81	0.80
(6)	**0.83**	**0.88**	**0.87**	**0.79**	**0.80**	**0.79**	**0.82**	**0.89**	**0.87**
(3)/(5)	9.8 %	4.4%	4.4%	7.2%	2.8 %	5.7 %	5.2%	3.8%	2.5%
(4)/(6)	12.1 %	23.9 %	22.5 %	6.7%	8.1 %	9.7 %	10.8 %	14.1 %	11.5 %

Table III. Number Of Adjectives And Verbs Implying Implicit Aspect For Each Social Media Review Dataset

	Social media review dataset 1	***Social media review dataset 2***	***Hate social media review dataset***
Number of sentences	357	641	648
Number of adjectives	406	841	773
Number of verbs	446	872	729

CONCLUSION

We proposed a hybrid technique that combines MNB, SVM, and RF classifiers to identify sentiment in social media reviews. The technique incorporated TF-IDF and WordNet Synonym relations of adjectives and verbs in document representation. We also proposed the best WN subsets of adjectives and verbs for document representation. The hybrid technique also included the use of a sentiment lexicon for sentiment analysis. This method was tested on several datasets and it achieved higher accuracy than the existing methods. The results showed that the hybrid technique is a promising approach for sentiment analysis in social media reviews.

REFERENCES

[1] M. Hu and B. Liu, "Mining and summarizing customer reviews," in Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, 2004, pp. 168–177.

[2] Doaa Mohey El-Din, "Enhancement Bag-of-Words Model for Solving the Challenges of

Sentiment Analysis" International Journal of Advanced Computer Science and Applications(IJACSA), 7(1), 2016.

[3] B. Liu, Sentiment analysis: mining opinions, sentiments, and emotions. New York, NY: Cambridge University Press, 2015.

[4] X. Chen, Y. Cho, and S. Y. Jang, "Social media review prediction using Twitter sentiment and weather," 2015, pp. 63–68.

[5] Jermy Prichard, Paul Watters, Tony KRONE, Caroline Spiranovic, and Helen Cockburn, "Social Media Sentiment Analysis: A New Empirical Tool for Assessing Public Opinion on Crime?," pp. 217–236, 2015.

[6] Nisal Waduge, "Machine Learning Approaches For Detect Crime Patterns - Data Gathering and Analysing Techniques," 2017.

[7] P. Burnap et al., "Detecting tension in online communities with computational Twitter analysis," Technol. Forecast. Soc. Change, vol. 95, pp. 96–108, Jun. 2015.

[8] N. Zainuddin, A. Selamat, and R. Ibrahim, "Improving Twitter Aspect- Based Sentiment Analysis Using Hybrid Approach," in Intelligent Information and Database Systems, vol. 9621, N. T. Nguyen, B. Trawiński, H. Fujita, and T.-P. Hong, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2016, pp. 151–160.

[9] Hissah AL-Saif and Hmood Al-Dossari, "Detecting and Classifying Social medias from Arabic Twitter Posts using Text Mining Techniques" International Journal of Advanced Computer Science and Applications(IJACSA), 9(10), 2018.

[10] B. Keith, E. Fuentes, and C. Meneses, "A Hybrid Approach for Sentiment Analysis Applied to Paper Reviews," N. S., p. 10, 2017.

[11] K. Schouten, P. O. Box, and D. Rotterdam, "Supervised and Unsupervised Aspect Category Detection for Sentiment Analysis with Co-occurrence Data," IEEE Trans. Cybern., p. 13, 2017.

[12] V. S. Jagtap and K. Pawar, "Sentence-Level Analysis of Sentiment Classification," Natl. Conf. Emerg. Trends Eng. Technol. Archit., p. 6, 2013.

[13] K. Gull, S. Padhye, and D. S. Jain, "A Comparative Analysis of Lexical/NLP Method with WEKA‟s Bayes Classifier," Int. J. Recent Innov. Trends Comput. Commun., vol. 5, no. 2, p. 7, 2017.

[14] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10, 2002, pp. 79–86.

[15] A. Alghunaim, M. Mohtarami, S. Cyphers, and J. Glass, "A Vector Space Approach for Aspect Based Sentiment Analysis," 2015, pp. 116– 122.

[16] G. Abdulsattar A. Jabbar Alkubaisi, S. Sakira Kamaruddin, and H. Husni, "Conceptual Framework for Stock Market Classification Model Using Sentiment Analysis on Twitter Based on Hybrid Naïve Bayes Classifiers," Int. J. Eng. Technol., vol. 7, no. 2.14, p. 57, Apr. 2018.

[17] V. N. Patodkar and S. I.R, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining," IJARCCE, vol. 5, no. 12, pp. 320–322, Dec. 2016.

[18] S. Rosenthal, N. Farra, and P. Nakov, “SemEval-2017 Task 4: Sentiment Analysis in Twitter,” p. 17, 2017.

[19] B. Gokulakrishnan, P. Priyanthan, T. Ragavan, N. Prasath, and As. Perera, “Opinion mining and sentiment analysis on a Twitter data stream,” 2012, pp. 182–188.

[20] M. Ishtiaq, “Sentiment Analysis of Twitter Data Using Sentiment Influencers,” vol. 6, no. 1, p. 9, 2015.

[21] H. Wang, D. Can, A. Kazemzadeh, F. Bar, and S. Narayanan, “A System for Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle,” p. 6, 2012.

[22] Y. Liu, H. T. Loh, and A. Sun, “Imbalanced text classification: A term weighting approach,” Expert Syst. Appl., vol. 36, no. 1, pp. 690–701, Jan. 2009.