Aspect Based Sentiment Analysis with Text Compression

Apoorva T.[1*], Pradeep N.[2]

[1*]Department of CSE, Bapuji Institute of Engineering and Technology, Davanagere, India
[2]Department of CSE, Bapuji Institute of Engineering and Technology, Davanagere, India

Corresponding Author: apoorvatammu@gmail.com

Abstract: Sentiment Analysis measures the aptitude of people's opinions through Natural Language Processing, Computational Lingus tics and Text analysis, which are used to extract and analyse subjectivity of information. This paper focuses on Aspect Based Sentiment Analysis, where Text Compression is performed before Aspect Based analysis. For a given huge text is compressed using Text compression model, which is considered as pre-processing task for Aspect Based Sentiment Analysis.

Keywords: Aspect Based sentiment analysis, text compression

I. INTRODUCTION

The internet holds a considerable amount of user-generated content describing the opinions of customers on products and services through blogs, tweets and other social media forms. Sentiment analysis and opinion mining have attracted significant attention in recent years as they pave the way for the automatic analysis of user reviews and the extraction of information most relevant to users [1], [2], [3-23].

Sentiment analysis entails several interesting and challenging tasks. One traditional and fundamental task is polarity classification, which determines the overall polarity (e.g., positive or negative) of a sentence or document However, these tasks are coarse-grained and cannot provide detailed information, such as the aspects on which the users comment. Recently, there has been a shift towards the fine-grained tasks, such as aspect-based (or "feature-based") sentiment analysis, which not only involves analysing the opinionated text's polarity (e.g., positive, neutral, negative) and intensity (e.g., weak, medium, strong, extreme), but also identifying the aspect (or the topic, or target entity) of the opinion [4]. The text compression performing before aspect based sentiment analysis as a sequence of labelling task, which can be solved using the Conditional Random Fields (CRF) model. Instead of seeking the manual rules on parse trees for compression, as in other studies [16], the CRF-based method is an automatic procedure.

II. RELEATED WORK

Earlier research on sentiment analysis primarily focused on polarity classification, i.e., determining the sentiment orientation of a sentence or a document [2]. However, these tasks are all coarse-grained and cannot provide more detailed information. Recently, there has been a shift towards fine-grained aspect-based tasks that can identify both the text expressing the opinion and the aspect of the opinion as well as analysing its polarity (e.g., positive, neutral or negative) [21], [22], [23]. The Aspect Polarity collocation extraction is the basic task of aspect-based sentiment analysis.

Tree-based approaches were commonly used to compress sentences which created a compressed sentence by making edits to the syntactic tree of the original sentence. However, the automatic parsing results may not be correct; thus, the compressed tree (after removing constituents from a bad parse) may not produce a suitable compressed sentence. To solve this problem using discriminative models, studies on classifiers to determine which words could be dropped by including features of the words themselves as well as part of the speech tags and parser trees. Here, the parser trees were soft evidence to determine whether to remove a word. Accordingly, the influencers of the parsing errors were reduced. Moreover, to improve the efficiency of the compression model, recent studies have been performed on polynomial time inference algorithms [40] and approximate inference algorithms [11] for sentence compression.

Currently, the existing text compression methods all focus on formal sentences, and few methods have been studied for sentiment sentences. The current compression models cannot be directly transplanted to sentiment sentences due to the specificity of the aspect based sentiment analysis. Therefore, a new compression model for sentiment sentences should be established.

The main goal of Aspect Based Sentiment Analysis is to identify the aspects of the given target entities and sentiment expressed for each aspect. To meet the objectives of Aspect

Based Sentiment Analysis, the work can be done primarily in two phases. In phase-1 extraction of aspect terms and grouping aspect terms into aspect categories will be done. In phase-2 identification of polarity of the aspect terms and polarity of the aspect categories of each sentence will be

done. These phases of activities are divided into four sub tasks namely: Aspect Term Extraction (ATE), Aspect Term Polarity (ATP), Aspect Category Detection (ACD) and Aspect Category Polarity (ACP).
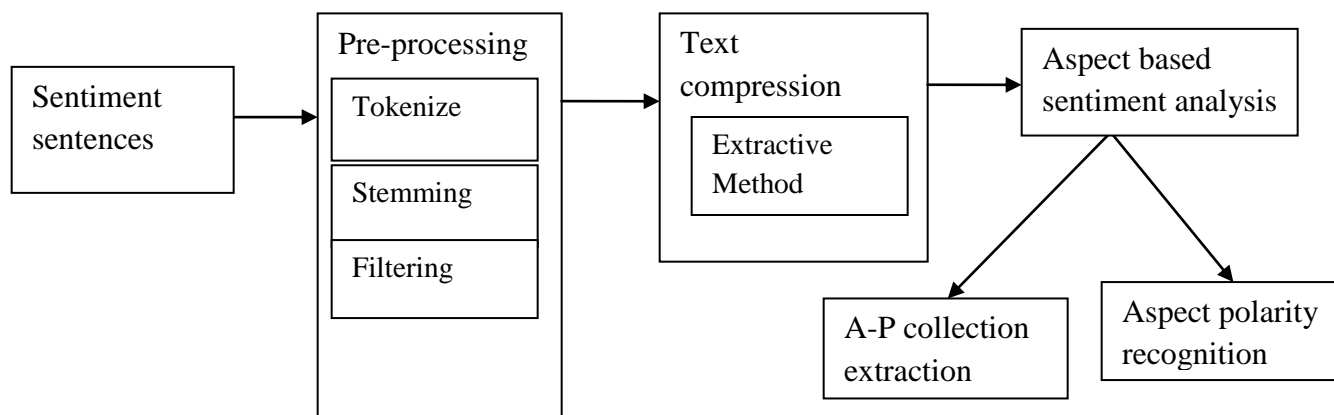
III METHODOLOGY

Figure 1: Framework of text compression model.

The pre-processing step for aspect based sentiment analysis has three steps, tokenization, stemming and filtering. Tokenization is the process of breaking a stream of text into words or phrases or other meaningful elements called tokens. This list of tokens becomes the input for further processing. Stemming is process of reducing inflected words to their word stem, base or root. Filtering involves removing repeated words that show the potential of expressing, which are not necessary for sentiment analysis [20]. The text compression is carried out using Extractive Method, because this method preserves the essential content of the text by dropping certain unimportant words from sentiment analysis.

The framework presents the double propagation method to extract the Aspect Polarity collocations, aspects and polarity words. This idea is based on the observation that there are natural syntactic relations between polarity words and aspects owing to the fact that polarity words are used to modify the aspects. Furthermore, they also discovered that the polarity words and aspects themselves had relations in certain sentiment sentences [22].

Various features used in text compression model are

- Basic Features.
- sentiment-related features
- potential semantic features
- syntactic parse features

Two types of sentiment-related features are Perception Features and Polarity Features. The perception feature indicates whether a word is a perception word. This type of feature is inspired by the naturalness problem. The current parser produces wrong parse trees because of these perception words [2]. Therefore, the perception words tend to be removed from a sentence from the text compression model. The polarity feature indicates whether a word is a polarity word.

IV. RESULTS

Words Count without Compression : 1234

Figure (a)

Without Text Compression Sentence polarity:

pos

Time taken for Sentiment Analysis without Text Comparession : 170.04

Figure (b)

Analysis with Text compression

Words Count with Compression: 452

With Compression Sentence polarity:
pos

Figure (c)

Time taken for Sentiment Analysis with Text Compression : 73.59

Figure (d)

Here we observe the results of the system, fig (a) shows the input file given for Sentiment analysis, fig (b) shows the total word count performed without text compression, fig (c) shows the text polarity performed without text compression and time taken for sentiment analysis without text compression, fig (d) shows the word count performed after text compression by removal of unnecessary words and also shows the text polarity, fig (e) shows the time taken for sentiment analysis with text compression.

IV. CONCLUSION

The text compression model is developed to perform before Aspect Based Sentiment Analysis. Which is different from the common sentence compression model, the text compression model not only compresses the redundancy in the sentences, but also needs to retain the polarity-related information to maintain the sentence original polarities. Thus, the over-natural and spontaneous sentences can be compressed into more formal and easier-to-parse sentences after using the text compression model. The text compression is performed to improve the performance of aspect based sentiment analysis. From the results we can observe that the analysis performed with text compression reduces the word count and also reduces the time taken as compared with the analysis without text compression.

V. REFERENCES

[1] Supriya B. Moralwar and Sachin N. Deshmukh, "*Different Approaches of Sentiment Analysis*", International Journal of Computer Sciences and Engineering, Vol.3, Issue.3, pp.160-165, 2015.

[2] U. Aggarwal, G. Aggarwal, "*Sentiment Analysis : A Survey*", International Journal of Computer Sciences and Engineering, Vol.5, Issue.5, pp.222-225, 2017.

[3] Reshma Bhonde1, Binita Bhagwat2, Sayali Ingulkar, "*Sentiment Analysis Based on Dictionary Approach*" International Journal of Emerging Engineering Research and Technology Vol.3, Issue.1, pp.51-55, 2015.

[4] C. Nanda, M. Dua, "*A Survey on Sentiment Analysis*", International Journal of Scientific Research in Computer Science and Engineering, Vol.5, Issue.2, pp.67-70, 2017.

[5] Jeonghee yi, Tetsuya, "*Sentiment analysizer: Extracting Sentiments abut a text with Natural Language processing techniques*", Department of Computer Science, university of texas, Austin, TX 78712, USA.

[6] Shabina Dhuria, "*Natural Language Processing Techniques for Sentimenyal analysis*", International Journal of Emerging Trends and Technology in computer science, Vol.1, Issue.2, July 2015.

[7] Navpreet Kaur, Mohit Kakkar, "*A balanced sentiment analysis approach with stemming porter for neutralized emotion weightage*", International Journal of Advanced Research in Computer and Communication Engineering Vol.4, Issue.10, October 2015.

[8] Norjihan Abd Ghani1 and Siti Syahidah Mohamad Kamal, "*A Sentiment-Based Filteration*", Proceedings of the 5th International Conference on Computing and Informatics, International Conference on Computer and Communication Technology (ICOCT) 2015 11-13 August, 2015 Istanbul, Turkey, University Utara Malaysia.

[9] Anais Collomb, Crina Costea, "*A Compariative study of sentiment analysis methods*", university of lyon, INSA-Lyon, F-69621 Villeurbanne, France.

[10] Hassan Saif,1 Miriam Fernandez,1 Yulan He,2 Harith Alani, "*On Stopwords, Filtering and Data Sparsity for Sentiment Analysis*", School of Engineering and Applied Science, Aston University, UK.

[11] T. Cohn and M. Lapata, "*An abstractive approach to sentence compression*", ACM Trans. Intell. Syst. Technol., Vol.4, pp. 1–35, 2013.

[12] Rehab Duwairi, Mahmoud El-Orfali, "*Effects of Preprocessing Strategies on Sentiment Analysis*", Journal of Information Science 1–14, 2013.

[13] Hemalatha, Dr. G. O Saradhi Varma, " *Preprocessing the Informal Text for efficient Sentiment Analysis*", International Jounal of Emerging And Technology in Computer Science Vol.1, Issue.2, 2012.

[14] G. Qiu, B. Liu, J. Bu, and C. Chen, "*Opinion word expansion and target extraction through double propagation*", Computer Linguist., vol. 37, no. 1, pp. 9–27, 2011.

[15] Dr.V.Maniraj, V.Krishnaveni, "*K-Nearest Neighbor Grouping in Excess of Semantically Secure Encryption Interactive Evidence*", International Journal of Computer Sciences and Engineering, Vol.4, Issue.4, pp.288-291, 2016.

[16] Vivek Narayanan, Ishan Arora , Arjun Bhatia, "*Fast and accurate sentiment classification using an enhanced Naive Bayes model*", Proceedings of the 49th annual meeting of the Association for Computational Linguistics: Human Language Technologies. Vol.1. 2011.

[17] B. Pang and L. Lee, "*Opinion mining and sentiment analysis*", Found. Trends Informatics Retr., Vol.2, no. 1–2, pp. 1–135, Jan. 2008.

[18] K. Knight and D and Marcu,M. Galley and K. McKeown, "*Lexicalized Markov grammars for sentence compression*", in Proceedings Human Language Technology: Conference North Amer. Chapter Associates for Computer Linguist.; Proc. Main Conf., Rochester, NY, USA, Apr. 2007, pp. 180–187,

[19] V. Hatzivassiloglou and K. R. McKeown. "*Predicting the semantic orientation of adjectives*", In Proceedings of the 35th ACL Conference, pages 174–181.

[20] T. Nomoto, "*Discriminative sentence compression with conditional random fields*", Information Process Management, Vol.43, pp. 1571–1587.

[21] J. Guo, W. Che, H. Wang, and T. Liu, "*Revisiting embedding features for simple semi-supervised learning*", in Proceedings EMNLP 14, 2014.

[22] J. Kaur, S.S. Sehra, S.K. Sehra, "*A Systematic Literature Review of Sentiment Analysis Techniques*", International Journal of Computer Sciences and Engineering, Vol.5, Issue.4, pp.22-28, 2017.

[23] Naveen Kumar Laskari, Suresh Kumar Sanampudi, "*Aspect Based Sentiment analysis*", IOSR Journal of Computer Engineering, Vol.18, Issue.2, pp. 24-28, 2016.