ASPECT-BASED SENTIMENT ANALYSIS OF ONLINE REVIEWS USING WORD AND PARAGRAPH VECTORS BY DEEP LEARNING METHODS

Zohreh Madhoushi [1], Abdul Razak Hamdan[2], Suhaila Zainudin[3]
[1,2,3]Center of Artificial Intelligence Technology
Faculty of Information Science and Technology
Universiti Kebangsaan Malaysia
Bangi, 43650, Selangor, Malaysia
Suhaila.zainudin@ukm.edu.my

Abstract: Aspect-based Sentiment Analysis (ABSA) aims to extract major aspects of an item or product and predict the polarity of each aspect from the reviews. The importance of developing ABSA system is due to easing the process of decision making for customers and also suppliers to monitor their consumers by providing a decomposed view of rated aspects. Previous methods tend to produce too many non-aspects and miss many of important ones. Domain specific models are often not practical for this task. Also limited works focus on implicit aspects and sentiments. Therefore the goal of this research is to develop an improved ABSA model that finds the most relevant aspect including implicit aspects with more accurate polarity estimation for each aspect in product reviews in different domains. We are going to take advantage of deep learning methods, which has gain significant interest recently. Word2Vec technique will be used for aspect extraction phase, and Doc2Vec technique will use for aspect polarity detection phase. Using these techniques, the system will outperform the previous ones in terms of the accuracy of aspect extraction and polarity prediction. Since deep learning methods do not use additional information such as parser and hand crafted features and sentiment lexicons, we hope to find more relevant aspects and more accurate polarity in multi domains for ABSA.

***Keywords*:** sentiment analysis, deep learning, opinion mining, word vectors, paragraph vectors.

INTRODUCTION (Heading Trebuchet MS, 11 font size, bold)

Generally, ABSA of linguistic expressions is a challenging task due to the ambiguity of natural language such as polysemy of words and complex syntactic structures of sentences. Although a great deal of work has been done in the research community and many systems have also been built, the problem is still far from being solved. Specifically this research focuses on the following problems.

First, several studies have been carried out using Language Rule Methods (Liu et al. 2015, Lal and Asnani 2014). These methods tend to produce too many non-aspects and miss some of the important aspects. Models that are based on frequent nouns missed low frequent aspects (Hu and Liu 2004). Models that try to find nouns that are related to the adjective, usually extract so many irrelevant aspect. This is because reviews are full of irrelevant information. For example in the review sentence below about a specific model of canon digital camera, knowing that 'fat' and 'short' are opinion words, the model take 'hand' and 'finger' as aspects which both are non-aspects.

- "I have fat hand and short fingers".

Models that choose the closest adjective to each noun (aspect) usually find it in a window of 5-6 words (Lek and Poo (2013)). Therefore the result is limited to the window size. For example in the sentence below, using window size of 6 we cannot extract 'image quality' as an aspect for 'excellent', because it is out of the window.

- Image quality of this camera which I bought it in a reasonable price is excellent.

Those models that use dependency parser heavily depend on parsing performance (Zhuang et al. 2006, Wu et al. 2009, Lizhen et al. 2014). Since online reviews are full of typo mistake, improper punctuation and grammar mistakes, parsers produce many errors. From the review, most attention has been paid to extracting explicit aspect and few of them focused on implicit aspect. Aspects that are explicitly mentioned as nouns or noun phrases in a sentence are known as explicit aspects, for example 'picture quality' in the sentence "The picture quality of this phone is great". Aspects that are not explicitly mentioned in a sentence but are implied, for example, 'price' in the sentence "This car is so expensive.", or 'size' in the sentence "This phone will not easily fit in a pocket".

Second, models that use closest adjective as sentiment word, cannot detect other sentiments which are not adjectives or implicit sentiments. For example in the previous sentence "this phone will not easily fit in a pocket", there is no adjective in the sentence but it implies negative opinion. Models that use supervised learning method (Smailović et al. 2014, Singh and Husain 2014, Da Silva et al. 2014) for polarity prediction in sentence level cannot predict the polarity of more than one aspect in a sentence. In another words they predict one polarity for all aspects in a sentence. Therefore the current F-measure results for different models range from 0.60 to 0.90, depending on domains and datasets. So the research in this area in not exhaustive yet, and needs improvement in term of accuracy of polarity prediction.

Finally, as observed from prior studies, it is difficult to obtain performance enhancement in different domains for supervised learning model. The reason is that a very large amount of labeled data in one domain is needed for the model to work properly. Also supervised models that develop in one domain usually do not work in another domain. It is also difficult to find domain dependent orientation of opinions in lexicon based models. For example quite is positive in restaurant domain but negative in MP3 player domain. Using lexicon to estimate the polarity of quiet, ignoring the domain, the model assign one polarity to both domains. Therefore, another motivation for this research is to target the problem of different domain.

The current F-measure results range from 0.60 to 0.90 depending on domains and datasets. Thus, the problems, remain to be highly challenging. We expect that this work improve the accuracy of aspect extraction and aspect polarity estimation. In this research we try to develop a model for ABSA of product review that work in different domain. Among three categories that we reviewed in the literature, language rule models, sequential models and topic models, we focus on language rule method. This is more relevant in this research due to sequential models are very much domain dependent, and topic models are too statistic centric that leads to limitation in NLP tasks that deal with meaning. Additionally, comparing these groups of methods a larger number of recent works use language rule models for ABSA tasks. Furthermore using deep learning tools in this research, hopefully we are going to be more close to the idea of creating a system that can understand meaning in text.

METHODS

Moghaddam and Ester (2010) states more than 70% of aspects are explicit noun or noun phrases. Hu and Liu (2004) assume that most aspects are noun phrases. This idea is used in many recent benchmark works (Lal and Asnani (2014), Moghaddam and Ester (2010)). Wu et al. (2009) observed that more than 98% of aspects are in noun phrase or verb phrase. We observe that aspects can be noun, adjective, adverb or verb including implicit and explicit aspects. In preprocessing phase we divide each review to sentences or sentence segments. Then we POS-tag all sentences. We use some seed words as aspect. To find aspects we want to use an unsupervised method (word2vec) which has not been used for the task of aspect extraction before. This technique shows considerable improvement in many NLP tasks such as text classification. And it does not need any additional information such as parse trees. This technique is in the family of deep learning methods which has gain significant interest recently. We will use Word2vec on our review dataset to create a vector representation of each word in the review. Using seed aspect in review websites, we will calculate the average of each category of these seed aspects. First we remove all general noun or noun phrases that contain, restaurant, cellphone, hotel and laptop. For each sentence we do the following procedure:

For each noun or noun phrases if its cosine similarity with any of aspects categories is more than a specific threshold then the candidate aspect is remain, otherwise it will be ignored. Cosine similarity is a measure of similarity between two vectors of an inner product space that measures the cosine of the angle between them. The cosine of two vectors can be derived by using the Euclidean dot product formula:

If no similar noun or noun phrase is found we do the same procedure for adjective. If no similar adjective is found we do the same procedure for adverb. And if no similar adverbs is found, we do the procedure for verb. This way we can find not only explicit aspects but some of the implicit ones as well. Figure 7 shows the process of aspect extraction.

For sentiment estimation, we follow the method introduce by Le and Mikolov (2014). They developed a deep learning model for sentiment analysis in paragraph and sentence level. But we want to use the model for ABSA. First we perform data preprocessing and create one sentence per line for all reviews in our dataset for both labeled and unlabeled dataset. Then we apply predefined rules to break the sentence where there are sentiment shifter words such as 'but' and 'however' for both labeled and unlabeled dataset. Then we will use Doc2Vec technique for unsupervised pre-training phase and logistic regression as the supervise phase of deep learning method, to predict the sentiment of each sentence or sentence segment. We assume the sentiment of each aspect is the result of the sentiment of those sentences or sentence segment that contain that aspect/group of aspect. We follow Hu and Liu

(2004) and require a polarity score for each occurrence of an aspect term. The polarity scores of all the occurrences of an aspect term, however, can then be averaged over all the input texts. If an aspect aggregation stage is also present, the polarity scores of aspect terms that have been clustered together can also be averaged. We are going to test the model on product review Multi-Domain Sentiment Dataset. We are going to evaluate our model from two different points of view: the accuracy of aspect extraction (feature selection), and the accuracy of sentiment estimation; and compare our model to the state of the art model for ABSA. We will utilize NLTK library in Python language for this research.

If the necessary ground truth is available, the performance of a method for aspect-based opinion mining can be evaluated by measures such as accuracy, precision and recall. However, in real-life data sets such ground truth is typically not available. In some of the works some human judges have been asked to read a set of reviews and manually create a set of "true" aspects and their polarities for the reviewed item as "gold standard". Precision and recall of aspect extraction are then computed versus this gold standard.

To evaluate performance of polarity prediction we can also use Precision, recall and F-measure as used in previous work shown in table 3. In our case, positive precision is the number of aspect term occurrences correctly classified as positive divided by the total number of aspect term occurrences classified as positive. Positive recall (R+) is the number of aspect term occurrences correctly classified as positive divided by the total number of truly positive aspect term occurrences. The precision and recall of the negative and neutral classes are defined similarly.

FINDINGS AND ARGUMENT

The importance of this study is not only due to easing the process of decision making for customers by providing a decomposed view of rated aspects (Figure 1), but also due to the ability of utilizing the extracted rated aspects in other opinion mining systems, for example opinion summarization systems and opinion question answering systems. Text summarization involves reducing a text document or a larger corpus of multiple documents into a short set of words or sentences that conveys the main meaning of the text (Kim et al. 2011; Meng et al. 2012; Nishikawa et al. 2010). Opinion QA methods try to answer questions using reviewers' opinions about target items (Moghaddam &Ester 2011), for example "Do people recommend digital camera X?". The results of ABSA can also be used in other computer systems, such as recommendation systems (to provide explanations for recommendation), advertising system (to place an ad of a product with similar rated aspects), and many business tasks related to sale management, reputation management, and public relations. This work will improve the state of the art of existing methods for ABSA. For each task of ABSA, the research will propose new methods (or improvement over previous methods), showing experimentally on dataset that the new methods (or the improved version) are better or at least comparable to state of the art ones in terms of relevant aspect extraction, accurate polarity prediction, and ability of the system to work in more than one domain. We will use distributed representation of words and document in this study. These techniques showed considerable improvement is sentiment analysis but to our knowledge did not use for ABSA before.

CONCLUSIONS

Although the field of opinion mining is new, but still diverse methods available to provide a way to do different tasks at different levels, with an outcome of innumerable applications. There should be a way to compare these techniques in different tasks at different levels. Since the nature of data set used varies in different work, existing evaluation metrics of different methods does not normally clarify the effectiveness of each method compare to others. Talking in general successful techniques are likely to be a good integration of hybrid approaches and natural language processing techniques. The open problems are that recent techniques are still unable to work well in different domain; sentiment classification based on insufficient labeled data is still a challenging problem; there is lack of SA research in languages other than English; and existing techniques are still unable to deal with complex sentences that requires more than sentiment words and simple parsing.

REFERENCES

Da Silva, N. F., Hruschka, E. R. & Hruschka Jr, E. R. (2014). Tweet Sentiment Analysis with Classifier Ensembles. *Decision Support Systems*.

Hu, M. & Liu, B. (2004). Mining and Summarizing Customer Reviews. *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. Pages 168-177.

Hu, X., Tang, L., Tang, J. & Liu, H. (2013). Exploiting Social Relations for Sentiment Analysis in Microblogging. Proceedings of the Sixth ACM International Conference on Web Search and Data Mining. Pages 537-546.

Kim, H. D., Ganesan, K., Sondhi, P. & Zhai, C. (2011). Comprehensive Review of Opinion Summarization. Retrieved from http://kavita-ganesan.com/mypublications.

Lal, M. & Asnani, K. (2014). Aspect Extraction & Segmentation in Opinion Mining. International Journal Of Engineering And Computer Science 3(5): 5873-5878.

Le, Q. V. & Mikolov, T. (2014). Distributed Representations of Sentences and Documents. arXiv preprint arXiv:1405.4053

Lek, H. H. & Poo, D. C. (2013). Aspect-Based Twitter Sentiment Classification. *Tools with Artificial Intelligence (ICTAI), 2013 IEEE 25th International Conference on*. Pages 366-373.

Liu, P., Qian, X. & Meng, H. (2015). Seeing Stars from Reviews by a Semantic-Based Approach with Mapreduce Implementation. 2014 ASE BigData/SocialInformatics/PASSAT/BioMedCom 2014 Conference.

Lizhen, L., Wei, S., Hanshi, W., Chuchu, L. & Jingli, L. (2014). A Novel Feature-Based Method for Sentiment Analysis of Chinese Product Reviews. *Communications*. 11(3), 154-164.

Long, C., Zhang, J. & Zhut, X. (2010). A Review Selection Approach for Accurate Feature Rating Estimation. *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*. Pages 766-774.

Meng, X., Wei, F., Liu, X., Zhou, M., Li, S. & Wang, H. (2012). Entity-Centric Topic-Oriented Opinion Summarization in Twitter. *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. pages 379-387.

Moghaddam, S. & Ester, M. (2010). Opinion Digger: An Unsupervised Opinion Miner from Unstructured Product Reviews. *Proceedings of the 19th ACM international conference on Information and knowledge management*. Pages 1825-1828.

Nishikawa, H., Hasegawa, T., Matsuo, Y. & Kikui, G. (2010). Opinion Summarization with Integer Linear Programming Formulation for Sentence Extraction and Ordering. *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*. Pages 910-918.

Qiu, G., Liu, B., Bu, J. & Chen, C. 2011. Opinion Word Expansion and Target Extraction through Double Propagation. *Computational Linguistics* , 37(1), 9-27.

Singh, P. K. & Husain, M. S. 2014. Methodological Study of Opinion Mining and Sentiment Analysis Techniques. *International Journal on Soft Computing*, 5(1).

Smailović, J., Grčar, M., Lavrač, N. & Žnidaršič, M. (2014). Stream-Based Active Learning for Sentiment Analysis in the Financial Domain. *Information Sciences*.

Wu, Y., Zhang, Q., Huang, X. & Wu, L. 2009. Phrase Dependency Parsing for Opinion Mining. Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3-Volume 3.Pages. 1533-1541.

Zhuang, L., Jing, F. & Zhu, X.-Y. (2006). Movie Review Mining and Summarization. *Proceedings of the 15th ACM international conference on Information and knowledge management*. pages 43-50.