Aspect-specific Sentimental Word Embedding for Sentiment Analysis of Online Reviews

Hui Du, Xueke Xu, Xueqi Cheng, Dayong Wu, Yue Liu, Zhihua Yu
Institute of Computing Technology, Chinese Academy of Sciences
duh@int-yt.com, xuxueke@software.ict.ac.cn, {cxq, wudayong, liuyue, yzh}@ict.ac.cn

ABSTRACT

Recently, Deep Convolutional Neural Networks (CNNs) have been widely applied to sentiment analysis of short texts. Naturally, word embedding techniques are used to learn continuous word representations for constructing sentence matrix as input to CNN. As for sentiment analysis of customer reviews, we argue that it is problematic to learn a single representation for a word while ignoring sentiment information and the discussed aspects. In this poster, we propose a novel word embedding model to learn sentimental word embedding given specific aspects by modeling both sentiment and syntactic context under the specific aspects. We apply our method as input to CNN for sentiment analysis in multiple domains. Experiments show that the CNN based on the proposed model can consistently achieve superior performance compared to CNN based on traditional word embedding method.

Categories and Subject Descriptors

I.5.4 [**PATTERN RECOGNITION**]: Applications – *Text processing*; I.2.7 [**ARTIFICIAL INTELLIGENCE**]: Natural Language Processing – *Language models*

General Terms

Algorithms, Experimentation

Keywords

word embedding; sentiment analysis; CNN; online reviews

1. INTRODUCTION

In this poster, we present a novel word embedding model for sentiment analysis of online customer reviews based on Deep Convolutional Neural Networks (CNN). CNN has recently achieved superior results for sentiment analysis of short texts compared with traditional bag-of-words methods [1,7]. Naturally, word embedding learning techniques, such as Continuous Bag-of-Words (CBOW) and Skip-Gram[6], could be used to learn continuous vector representations of words for constructing sentence matrix as input to CNN. The word vectors can capture both syntactic and semantic information of the words by modeling surrounding contexts.

As for sentiment analysis of customer reviews, traditional word embedding techniques should be improved in the following two points. First, they usually only model syntactic contexts of the word while ignoring its sentiment behaviors. Two words sharing similar syntactic contexts may have opposite sentiment polarities. Second, and more importantly, learning a single vector for a word is problematic since the sentiment polarity of a word is usually sensitive to the target aspects. Indeed, customers usually talk about different aspects of the object in reviews, such as *service* or *atmosphere* of a hotel. One single word may deliver different sentiment polarities according to the targeted aspects. For instance, for a hotel we may enjoy a *large* room, but not expect *large* noise.

We propose a novel word embedding model, called aspect-specific sentiment word embedding (ASWE), to learn sentiment-bearing word representations with respect to the specific aspects by modeling both sentiment and syntactic contexts under specific aspects. The proposed model is built on the basis of CBOW. It aims to learn an appropriate representation for each word-aspect pair that captures the sentiment of the word with respect to the aspect. To this end, each word token is beforehand assigned an aspect by aspect-extracting topic models. We consider three kinds of contexts, i.e., the surrounding word-aspect pairs, the word itself, and the sentiment label, to jointly predict the target word-aspect pair in a parallel way.

The word-aspect vectors learned by our method are used as input to CNN for sentiment analysis of reviews. Experiments in multiple domains show that the CNN based on ASWE can consistently achieve superior performance compared to CNN based on traditional word embedding methods.

2. The Proposed Model

NOTATIONS: Before presenting our model, we first present the notations used in this paper. Consider a set of aspect labels $\boldsymbol{A}$ and a set of sentiment labels $\boldsymbol{S}$={**negative, positive**}. Let D $\{d \ \ldots\ d_N\}$ denotes a corpus of *N* reviews. Each review $d_m \in D$ consists of a sequence of T_m words from the vocabulary V, where each word w_i^m is assigned an aspect label $a_i^m \in \boldsymbol{A}$ as the aspect in discussion. Aspect-extracting topic modeling methods can be used to assign aspects to word tokens. Then, each document is treated as a sequence of word-aspect pairs, i.e., $d_m \ (\ w^m \ a^m \ \cdots \ w_i^m \ a_i^m \ \cdots \ w_{T_m}^m \ a_{T_m}^m \ )$. Each review d_m is also associated with a sentiment label $s_m \in \boldsymbol{S}$. Review sentiment labels are easily available since each review is usually associated with a customer rating.

In this model, we will learn representation vectors for words with respect to aspects (i.e., word-aspect pairs), words regardless of any aspects, as well as sentiment labels in a common space. Note that, when training and testing with CNN, for each word-aspect pair, the average of the learned vectors for the word-aspect pair and the word will be used for constructing input matrix.

The architecture of the model can be shown in Figure 1. Each target word-aspect pair is predicted by its surrounding word-aspect pairs, as well as the word itself and the sentiment label of the review where the word occurs in a parallel way. More formally, the objective of the model is to maximize the prediction probabilities of all word-aspect pairs in the corpus, and the objective function consists of the following three parts:

WWW 2016 Companion, April 11-15, 2016, Montréal, Québec, Canada.
ACM 978-1-4503-4144-8/16/04.
http://dx.doi.org/10.1145/2872518.2889403

$$\ell = \sum_{m=1}^{N} \sum_{i=1}^{T_m} (\log p(< w_i^m, a_i^m > | h_i^m) + \log p(< w_i^m, a_i^m > | s_m) + \log p(< w_i^m, a_i^m > | w_i^m))$$

Where h_i^m means the projection of surrounding context of the target word, i.e.,

$$h_i^m = avg(< w_{i-L}^m, a_{i-L}^m >, ..., < w_{i-1}^m, a_{i-1}^m >, < w_{i+1}^m, a_{i+1}^m > ..., < w_{i+L}^m, a_{i+L}^m >)$$

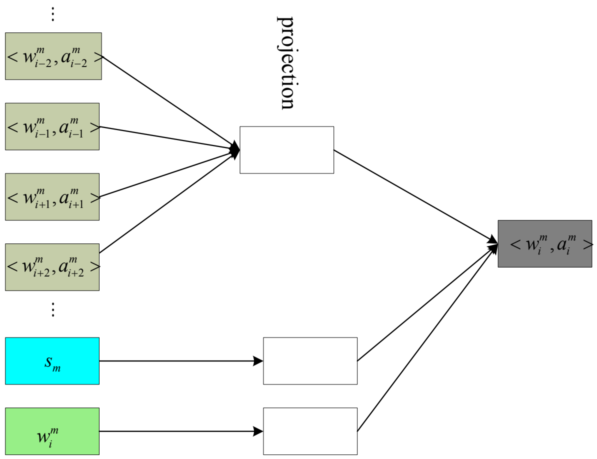

Figure 1. The framework of ASWE model.

The first part captures aspect-specific semantic information of the word by modeling words-by-surrounding-contexts co-occurrences under the specific aspects. While the second part captures the aspect-specific sentiment information by modeling words-by-sentiments co-occurrences under the specific aspects. We could imagine each sentiment as a large virtual document consisting of all word-aspect pairs from all reviews with the specific sentiment label. It is notable that the third part captures the common semantic information of words across different aspects. This part is necessary because dividing words by aspects will somewhat suffers from sparseness and wrong aspect assignments, especially for long-tailed words.

Note that, [3] proposed a topical word embedding model for computing similarity between words with respect to the specific topical contexts. In two of three variants they presented, no topic-specific word representations is learned, but instead, additional topic embedding is concatenated with word embedding for computing contextual similarity. The topic embedding may capture the general semantic information of the topic, without considering the special sentiment information of individual words. Another variant learns topic-specific word embedding by treating word-topic pairs as words. But experiment results were not desirable, largely due to the sparse problem. Furthermore, no sentiment information is considered in all of the three variants.

3. EXPERIMENTS

We use a large scale of review data[1] collected from *Amazon* by McAuley [4, 5]. The data is divided into different domains. Here we choose four most popular domains, i.e., *Electronics*, *Movies and TV*, *CDs and Vinyl* and *Clothing, Shoes and Jewelry*, with enough reviews for word embedding training. Each review is associated with a rating by the user on a scale of 5, where 4 and 5 stars means positive, and 1 and 2 stars means negative. We here ignore reviews with 3 stars since the sentiment is ambiguous. For each domain, we randomly choose 99% reviews for word embedding training. Within the remaining reviews, we randomly choose half for CNN training and the rest for testing. As for aspect assignments for word tokens, we choose Sentence-LDA (SLDA) [2]. SLDA is an extension to LDA by constraining that all words of each sentence are assigned to a single topic to make the extracted topics correspond to the reviewable aspects. We perform collapsed Gibbs sampling to iteratively assign aspect for each word token, and then choose the aspect with max assignment times.

We implement our model based on the code of CBOW in word2vec [6], and apply negative sampling technique for learning. We adopt default parameters as word2vec, except that the dimensionality is set to 50. Our CNN framework is most similar to [7] which has recently yielded state-of-the-art results. The major differences lie in that we use multiple windows sizes, each size has a set of filters, and that we use tanh function as activation units.

The final results are shown in table 1 in terms of accuracy. Here, SentiCBOW is an extension to CBOW by predicting the target word with both the surrounding context and the sentiment label, but without considering any aspects. As comparison, we use standard CBOW and SentiCBOW to train word embedding as input to CNN. As the table shows, ASWE can consistently achieve superior performance compared with traditional word embedding method (CBOW) and aspect-unware method (SentiBOW).

Table 1. Experiments results in four domains

Model	***Electronics***	***Movies and TV***	***CDs and Vinyl***	***Clothing, Shoes and Jewelry***
CBOW	90.41%	91.31%	91.95%	90.49%
SentiCBOW	91.12%	90.15%	92.26%	92.03%
ASWE	**92.08%**	**92.05%**	**94.38%**	**93.22%**

4. ACKNOWLEDGMENTS

This work is funded by 973 Program of China (Nos. 2014CB340401 and 2012CB316303), 863 Program of China (Nos. 2014AA015204 and 2014AA015103), the NSFC (Nos. 61232010 and 61425016), and Key Research Program of the Chinese Academy of Sciences under Grant NO.KGZD-EW-T03-2.

5. REFERENCES

[1] Nal Kalchbrenner, Edward Grefenstette and Phil Blunsom. 2014. A Convolutional Neural Network for Modelling Sentences. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics.*

[2] Yohan Jo and Alice H. Oh. 2011. Aspect and sentiment unification model for online review analysis. In *Proceedings of the fourth ACM international conference on Web search and data mining* (WSDM '11).

[3] Yang Liu, Zhiyuan Liu, Tat-Seng Chua, Maosong Sun.2015. Topical Word Embeddings. In *AAAI'15*.

[4] J. McAuley, R. Pandey, J. Leskovec.2015. Inferring networks of substitutable and complementary products. *Knowledge Discovery and Data Mining*, 2015

[5] J. McAuley, C. Targett, J. Shi, A. van den Hengel.2015. Image-based recommendations on styles and substitutes. In *SIGIR '15*.

[6] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. 2013. Distributed representations of words and phrases and their compositionality. In *NIPS'13*.

[7] Aliaksei Severyn and Alessandro Moschitti. 2015. Twitter Sentiment Analysis with Deep Convolutional Neural Networks. In *SIGIR '15*.

[1] . http://jmcauley.ucsd.edu/data/amazon/links.html