International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT

2020 IJIEMR.Personal use of this material is permitted. Permission from IJIEMR must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. No Reprint should be done to this paper, all copy right is authenticated to Paper Authors

IJIEMR Transactions, online available on 2nd Jan 2021. Link

:http://www.ijiemr.org/downloads.php?vol=Volume-09&issue=ISSUE-12

DOI: 10.48047/IJIEMR/V09/I12/150

Title: **WEAKLY-SUPERVISED DEEP EMBEDDING FOR PRODUCT REVIEW SENTIMENT ANALYSIS**

Volume 09, Issue 12, Pages: 873-879

Paper Authors

SINDE TEJASHWINI, ABBURI ANUSHA, PENDLI ANUSHA, T.KANAKAIAH

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

WEAKLY-SUPERVISED DEEP EMBEDDING FOR PRODUCT REVIEW SENTIMENT ANALYSIS

SINDE TEJASHWINI[1], ABBURI ANUSHA[2], PENDLI ANUSHA[3], T.KANAKAIAH[4]

[1,2,3,] B TECH Students, Department of CSE, Princeton Institute of Engineering & Technology For Women, Hyderabad, Telangana, India.

[4] Assistant Professor, Department of CSE, Princeton Institute of Engineering & Technology For Women, Hyderabad, Telangana, India.

Abstract: Sentiment analysis is one of the key challenges for mining online user generated content. In this work, we focus on customer reviews which are an important form of opinionated content. The goal is to identify each sentence's semantic orientation (e.g. positive or negative) of a review. Traditional sentiment classification methods often involve substantial human efforts, e.g. lexicon construction, feature engineering. In recent years, deep learning has emerged as an effective means for solving sentiment classification problems. A neural network intrinsically learns a useful representation automatically without human efforts. However, the success of deep learning highly relies on the availability of large-scale training data. In this paper, we propose a novel deep learning framework for review sentiment classification which employs prevalently available ratings as weak supervision signals. The framework consists of two steps: (1) learn a high level representation (embedding space) which captures the general sentiment distribution of sentences through rating information; (2) add a classification layer on top of the embedding layer and use labeled sentences for supervised fine-tuning. Experiments on review data obtained from Amazon show the efficacy of our method and its superiority over baseline methods.

I. INTRODUCTION

With the booming of Web 2.0 and e-commerce, more and more people start consuming online and leave comments about their purchase experiences on merchant/review Websites. These opinionated contents are valuable resources both to future customers for decision-making and to merchants for improving their products and/or service. However, as the volume of reviews grows rapidly, people have to face a severe information overload problem. To alleviate this problem, many opinion mining techniques have been proposed, e.g. opinion summarization [Hu and Liu, 2004; Ding et al., 2008], comparative analysis [Liu et al., 2005] and opinion polling [Zhu et al., 2011]. A key component for these opinion mining techniques is a sentiment classifier for natural sentences. Popular sentiment classification methods generally fall into two categories: (1) lexicon-based methods and (2) machine learning methods. Lexicon-based methods [Turney, 2002; Hu and Liu, 2004; Ding et al., 2008] typically take the tack of first constructing a sentiment lexicon of opinion words (e.g. "good", "bad"), and then design classification rules based on appeared opinion words and prior syntactic knowledge. Despite effectiveness, this kind of methods require substantial efforts in lexicon construction and rule design. Furthermore, lexicon-based methods cannot well handle implicit opinions, i.e. objective statements such as "I bought the mattress a week ago, and a valley appeared today". As pointed out in [Feldman, 2013], this is also an important form of opinions. Factual information is usually more helpful than subjective feelings. Lexicon-based methods can only deal with implicit opinions in an ad-hoc way [Zhang and Liu, 2011]. A pioneering work [Pang et al., 2002] for machine learning based sentiment classification applied standard machine learning algorithms (e.g. Support Vector Machines) to the problem. After that, most research in this direction revolved around feature engineering for better classification performance. Different kinds of features have been explored, e.g. n-grams [Dave et al., 2003], Part-of-speech (POS) information and syntactic relations [Mullen and Collier, 2004], etc. Feature engineering also costs a lot of human efforts, and a feature set suitable for one domain may not generate good performance for other domains [Pang and Lee, 2008]. In recent years, deep learning has emerged as an effective means for solving sentiment classification problems [Glorot et al., 2011; Kim, 2014; Tang et al., 2015; Socher et al., 2011; 2013]. A deep neural network intrinsically learns a high level representation of the data [Bengio et al., 2013], thus avoiding laborious work such as feature engineering. A second advantage is that deep models have exponentially stronger expressive power than shallow models. However, the success of deep learning heavily relies on the availability of large-scale training data [Bengio et al., 2013; Bengio, 2009]. Constructing large-scale labeled training datasets for sentence level sentiment classification is still very laborious. Fortunately, most merchant/review Websites allow customers to summarize their opinions by an overall rating score (typically in 5-stars scale). Ratings reflect the overall senti ment of customer reviews and have already been exploited for sentiment analysis [Maas et al., 2011; Qu et al., 2012]. Nevertheless, review ratings are not reliable labels for the constituent sentences, e.g. a 5-stars review can contain negative sentences and we may also see positive words occasionally in 1-star reviews. An example is shown in Figure 1. Therefore, treating binarized ratings as sentiment labels could confuse a sentiment classifier for review sentences. In this work, we propose a novel deep learning framework for review sentence sentiment classification. The framework leverages weak supervision signals provided by review ratings to train deep neural networks. For example, with 5- stars scale we can deem ratings above/below 3-stars as positive/negative weak labels respectively. It consists of two steps. In the first step, rather than predicting sentiment labels directly, we try to learn an embedding space (a high level layer in the neural network) which reflects the general sentiment distribution of sentences, from a large number of weakly labeled sentences. That is, we force sentences with the same weak labels to be near each other, while sentences with different weak labels are kept away from one another. To reduce the impact of sentences with rating-inconsistent orientation (hereafter called wrong-labeled sentences), we propose to penalize the relative distances among sentences in the embedding space through a ranking loss. In the second step, a classification layer is added on top of the embedding layer, and we use labeled sentences to fine-tune the deep network. Regarding the network, we adopt Convolutional Neural Network (CNN) as the basis structure since it achieved good performance for sentence sentiment classification [Kim, 2014]. We further customize it by taking aspect information (e.g. screen of cell phones) as an additional context input. The framework is dubbed Weakly-supervised Deep Embedding (WDE). Although we adopt CNN in this paper, WDE also has the potential to work with other types of neural networks. To verify the effectiveness of WDE, we collect reviews from Amazon.com to form a weakly labeled set of 1.1M sentences and a manually labeled set of 11,754 sentences. Experimental results show that WDE is effective and outperforms baselines methods.

II. RELATED WORK

Sentiment analysis is a long standing research topic. Readers can refer to [Liu, 2012] for a recent survey. Sentiment classification is one of the key tasks in sentiment analysis and can be roughly categorized as document level, sentence level and aspect level. Our work falls into the last category since we consider aspect information. In the next we review two subtopics closely related to our work.

Deep Learning for Sentiment Classification: In recent years, deep learning has received more and more attention in the sentiment analysis community. Researchers have explored different deep models for sentiment classification. Glorot et al. used stacked denoising auto-encoder to train review representation in an unsupervised fashion, in order to address the domain adaptation problem of sentiment classification [Glorot et al., 2011]. Socher et al. [Socher et al., 2011; 2012; 2013] proposed a series of Recursive Neural Network (RecNN) models for sentiment classification. These methods learn vector representations of variable-length sentences through compositional computation recursively. Kim investigated using CNN for sentence sentiment classification and found it outperformed RecNN [Kim, 2014]. A variant CNN with dynamic k-max pooling and multiple convolutional layers was proposed in [Kalchbrenner et al., 2014]. Researchers have also investigated using sequential models such as Recurrent Neural Network (RNN) and Long ShortTerm Memory (LSTM) for sentiment classification [Tang et al., 2015]. However, none of the above works tried to use review ratings to train deep sentiment classifiers for sentences. This is not a trivial problem since ratings are too noisy to be used directly as sentence labels (see Section 3 and experiments for discussions of this issue). To our knowledge, The WDE framework is the first attempt to make use of rating information for training deep sentence sentiment classifiers. Note that although we choose CNN as the deep model due to its competitive performance on sentiment classification [Kim, 2014], the idea of WDE could also be applied to other types of deep models. The major contribution of this work is a weakly supervised deep learning framework, rather than specific deep models.

Exploiting Ratings in Sentiment Classification: Rating information has been exploited in sentiment classification. Qu et al. incorporated ratings as weak labels in a probabilistic framework for sentence level sentiment classification [Qu et al., 2012]. However, their method still required careful feature design and relied on base predictors. While our method automatically learns a meaningful sentence representation for sentiment classification. Tackstr ¨ om and McDonald ¨ used conditional random fields to combine review level and sentence level sentiment labels for sentence sentiment analysis.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

This method also required feature engineering. Maas et al. [Maas et al., 2011] proposed to learn sentiment-bearing word vectors by incorporating rating information in a probabilistic model. For sentiment classification, they simply averaged the word vectors of a document as its representation. A similar work is [Tang et al., 2014], which developed a variant of the C&W neural model [Collobert et al., 2011] for learning sentiment-bearing word vectors from weak tweet labels derived from emoticons. The tweet representation was obtained by min, max and avg pooling on word vectors. Although this kind of methods can generate sentence representations automatically, the representations were derived by simple pooling of the learned word vectors. In comparison, our method generates a sentence representation by feeding word vectors through an expressive deep neural network. Moreover, we directly optimize sentence representation, rather than word vectors. We take the above two methods as baselines in experiments.

III. METHODOLOGY

Weakly-supervised Deep Embedding: The classic deep learning methods take an "unsupervised training then supervised fine-tuning" scheme, where restricted Boltzmann machines (RBM) or auto-encoders are used to pre-train network parameters from large quantities of unlabeled data [Bengio, 2009]. This works well when the data distribution is correlated with label prediction [Bengio, 2009]. Nevertheless, in sentiment analysis the word co-occurrence information is usually not well correlated with sentiment prediction [Maas et al., 2011], which motivates us to exploit large-scale rating data for training deep sentiment classifiers.

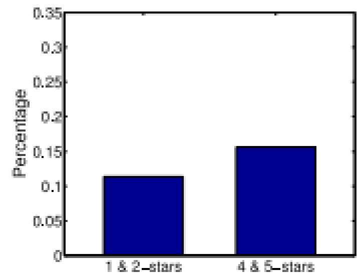

Figure 1: Percentages of wrong-labeled sentences by ratings in our labeled review dataset. The overall percentage is 13.4%.

However, ratings are noisy labels for review sentences and would mislead classifier training if directly used in supervised training. In this paper, we adopt a simple rule to assign weak labels to sentences with 5-stars rating scale: `(s) = ⇢ pos, if s is in a 4 or 5-stars review neg, if s is in a 1 or 2-stars review , (1) where `(s) denotes the weak sentiment label of sentence s. Figure 2 shows the percentages of wrong-labeled sentences by `(s), estimated in our labeled review dataset (detailed description of the dataset is in Section 4.1). We can see the noise level is moderate but not ignorable. The general idea behind WDE is that we use large quantities of weakly labeled sentences to train a good embedding space so that a linear classifier would suffice to accurately make sentiment predictions. Here good embedding means in the space sentences with the same sentiment labels are close to one another, while those with different labels are kept away from each other. In the following, we first present the network architecture, and then discuss how to train it with largescale rating data, followed by supervised fine-tuning on labeled sentences.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Network Architecture: The network architecture, depicted in Figure 3, is a variant of the CNNs described in [Collobert et al., 2011; Kim, 2014]. In what follows, we use upper case bold letters such as W to denote matrices and lower case bold letters such as x to denote column vectors. The i-th element in vector x is denoted by x(i).

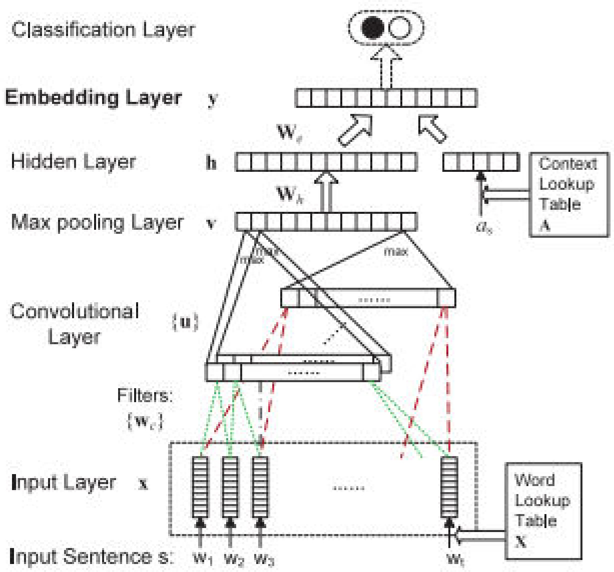

Figure 2: The CNN network architecture for sentence sentiment classification.

Input Layer: An input sentence of length t is a word sequence s =< w1w2 ...wt >. Each word w in the vocabulary is described by a word vector x. Let k be the length of x and n be the total number of words in the vocabulary. The trainable word lookup table X is then a k ⇥ n matrix with word vectors as its columns. The input layer simply maps s =< w1w2 ...wt > to its corresponding word vector representation < x1x2 ... xt >. The lookup table is initialized using the publicly available 300-dimensional word vectors trained on 100 billion words from Google News by word2vec [Mikolov et al., 2013]. Out-of-sample words are initialized randomly. **Convolutional Layer and Max pooling Layer:** The convolutional layer applies a set of filters on the sentence. Each filter w 2 Rhk is applied to a window of h words to produce a local feature value.

This pooling scheme keeps the most important indicator of a feature and naturally leads to a fixed-length vector output v at the max pooling layer. A filter with window size h is intrinsically a feature extractor which performs "feature selection" from the h-gram features of a sentence. When the input h-gram matches its w, we will obtain a high feature value, indicating this h-gram activates the feature. This resembles the traditional feature selection in sentiment classification [Pang and Lee, 2008], but is done automatically by the network. Since traditional machine learning based methods often exploit unigrams, bigrams and trigrams [Pang and Lee, 2008], we also employ filters with different window sizes, i.e. h = 1, 2, 3.

Hidden Layer and Embedding Layer: The fixed-length feature vector v is then fed to the fully connected hidden layer and embedding layer to extract nonlinear higher level features.

The embedding layer gets its input from two sources: the output of the hidden layer h, and context vector as of sentence s. A context vector is the semantic representation of an aspect that customers can comment on with respect to a sort of entities. For instance, battery life is an aspect for cell phones. The motivation for incorporating aspect information as the context of a sentence is that similar comments in different contexts could be of opposite orientations, e.g. "the screen is big" vs. "the size is big". Context vectors of all aspects constitute the context lookup table A (as columns).

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Supervised Fine-tuning: After obtaining a good enough sentence representation by the embedding layer, we add a classification layer on the top (Figure 3) to further train the network using labeled sentences. The classification layer simply performs standard affine transformation of the embedding layer output y and then applies a softmax activation function [Bishop, 2006] to the result for label prediction. In this work, we focus on binary sentiment prediction (i.e. positive or negative) since we only consider sentences which comment on specific aspects of an entity. These kinds of sentences hardly contain neutral sentences. Nevertheless, WDE could also be adapted to multi-class prediction problems. For binary prediction, the classification layer is equivalent to a logistic regression model. We train the network using standard SGD, since AdaGrad can easily "forget" the prior model learned in the first phase.

Varying the Size of Training Set: Next we examine the impact of the size of labeled training data on each method's performance. CNN-weak and Lexicon are not involved since they do not depend on labeled training data. We randomly select d% training data to train the classifiers and test them on the test set, with d ranging from 10 to 90. For each d, we generate the training set 30 times and the averaged performance is reported. Figure 5 shows the results. We can see that as the number of available training instances decreases, the performance of CNNrand, NBSVM and SVM drops faster than that of WDE, SSWE and SentiWV. This should be because the latter methods have gained prior knowledge about the sentiment distribution through pre-training, though with different capabilities. With 10% training set (nearly 600 instances), WDE can still achieve around 80% accuracy on the test set. According to t-test, WDE significantly outperforms the other methods with p-value < 0.01.

CONCLUSIONS

In this work we proposed a novel deep learning framework named Weakly-supervised Deep Embedding for review sentence sentiment classification. WDE trains deep neural networks by exploiting rating information of reviews which is prevalently available on many merchant/review Websites. The training is a 2-step procedure: first we learn an embedding space which tries to capture the sentiment distribution of sentences by penalizing relative distances among sentences according to weak labels inferred from ratings; then a softmax classifier is added on top of the embedding layer and we finetune the network by labeled data. Experiments on reviews collected from Amazon.com show that WDE is effective and outperforms baseline methods. For future work, we will investigate applying WDE on other types of deep networks and other problems involving weak labels.

REFERENCES

1. Yoshua Bengio, Aaron Courville, and Pierre Vincent. Representation learning: A review and new perspectives. IEEE TPAMI, 35(8):1798–1828, 2013.
2. Yoshua Bengio. Learning deep architectures for ai. Foundations and trendsR in Machine Learning, 2(1):1–127, 2009.
3. Christopher M Bishop. Pattern recognition and machine learning. springer, 2006.
4. Ronan Collobert, Jason Weston, Leon Bot- ´ tou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. Natural language processing (almost) from scratch. JMLR, 12:2493– 2537, 2011.

5. Kushal Dave, Steve Lawrence, and David M Pennock. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In WWW, pages 519– 528, 2003.
6. Xiaowen Ding, Bing Liu, and Philip S Yu. A holistic lexicon-based approach to opinion mining. In WSDM, pages 231–240, 2008.
7. John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. JMLR, 12:2121–2159, 2011.
8. Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. Liblinear: A library for large linear classification. JMLR, 9:1871–1874, 2008.
9. Ronen Feldman. Techniques and applications for sentiment analysis. Communications of the ACM, 56(4):82–89, 2013.
10. Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Domain adaptation for large-scale sentiment classification: A deep learning approach. In ICML, pages 513–520, 2011.
11. Minqing Hu and Bing Liu. Mining and summarizing customer reviews. In SIGKDD, pages 168–177, 2004.
12. Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. A convolutional neural network for modelling sentences. In ACL, 2014.
13. Yoon Kim. Convolutional neural networks for sentence classification. In EMNLP, pages 1746–1751, 2014.
14. Bing Liu, Minqing Hu, and Junsheng Cheng. Opinion observer: analyzing and comparing opinions on the web. In WWW, pages 342–351, 2005.
15. Bing Liu. Sentiment analysis and opinion mining. Morgan & Claypool Publishers, 2012.
16. Andrew L Maas, Raymond E Daly, Peter T Pham, Dan Huang, Andrew Y Ng, and Christopher Potts. Learning word vectors for sentiment analysis. In ACL, pages 142–150, 2011.
17. Julian McAuley, Rahul Pandey, and Jure Leskovec. Inferring networks of substitutable and complementary products. In SIGKDD, pages 785–794, 2015.
18. Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In NIPS, pages 3111–3119, 2013.
19. [Mullen and Collier, 2004] Tony Mullen and Nigel Collier. Sentiment analysis using support vector machines with diverse information sources. In EMNLP, volume 4, pages 412–418, 2004.
20. Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. Foundations and trends in information retrieval, 2(1-2):1–135, 2008.
21. Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up?: sentiment classification using machine learning techniques. In EMNLP, pages 79–86, 2002.
22. Lizhen Qu, Rainer Gemulla, and Gerhard Weikum. A weakly supervised model for sentence-level semantic orientation analysis with multiple experts. In EMNLP-CoNLL, pages 149–159, 2012.
23. Richard Socher, Jeffrey Pennington, Eric H Huang, Andrew Y Ng, and Christopher D Manning. Semisupervised recursive autoencoders for predicting sentiment distributions. In EMNLP, pages 151–161, 2011.