Juni Khyat
(UGC Care Group I Listed Journal)
ISSN: 2278-4632
Vol-11 Issue-01 2021

WEAKLY-SUPERVISED DEEP LEARNING FOR CUSTOMER REVIEW SENTIMENT CLASSIFICATION

SAATHWIK CHANDAN NUNE IIM KOZHIKODE

ABSTARCT:

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.

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

ISSN: 2278-4632
Vol-11 Issue-01 2021

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 sentiment 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.

EXISTING SYSTEM:

Lexicon-based methods typically take the tack of first constructing a sentiment lexicon of opinion words (e.g. “wonderful”, “disgusting”), and then design classification rules based on appeared opinion words and prior syntactic knowledge. Despite effectiveness, this kind of methods requires 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 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.

DISADVANTAGES:

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. This kind of algorithm needs complex lexicon construction and rule design. The existing systems cannot well handle objective statements; it only handles single word based sentiment analysis.

PROPOSED SYSTEM:

In this work, we propose a novel deep learning framework for review sentence sentiment classification. The framework treats review ratings as weak labels 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. The framework generally 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

ISSN: 2278-4632
Vol-11 Issue-01 2021

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. The framework is dubbed Weakly-supervised Deep Embedding (WDE). Regarding network structure, two popular schemes are adopted to learn to extract fixed-length feature vectors from review sentences, namely, convolutional feature extractors and Long Short-Term Memory.

ADVANTAGES:

The Proposed work leverages the vast amount of weakly labeled review sentences for sentiment analysis. It is much more effective than the previously developed works. The proposed work finds the sentiment not only based on the rating that user gives but also taking into consideration of reviews that they are post, In fact mainly takes an account of review, even though user gave ratings.

MODULES:

There are five modules divided in this project in order develop the concept of sentiment analysis with tagging. They are listed below

1. Products Initiation
2. Products acquisition
3. Sentiment classification
4. Weak Supervision
5. Graphical Analysis

MODULE DESCRIPTION:

Products Initiation

The First phase of the implementation of this project is Products Initiation. In this module admin is uploading the products which user wants to see and purchase. Once admin uploads the product means it stored in the database. The products which are uploaded are listed in website to admin in order to modify or delete the particular product. Admin is the only authorized person to upload the products in this project.

Products acquisition

The second module of this product conveys that user can view the products which are uploaded by admin. Then they can view the ratings and reviews of the same products which are given by other users who already purchased the product. According to the help of ratings and reviews user can purchase the product. The ordered list is also shown in the project for the convenience of users. The cart and checkout facility is also available to users from this module.

Sentiment classification

The users who are all purchased the products can rate product as per their interest on one scale of five and they are free to comment for the same. Based on the ratings and reviews given by user sentiment can be analyzed. There are two sentiments maintained in this project they are positive and negative. The equilibrium of rating and the particular comments are noted. In this module of project

ISSN: 2278-4632
Vol-11 Issue-01 2021

we implement the algorithm named Sentiment-Analysis-using-Naive-Bayes-Classifier to find the exact sentiment based on the dataset which are predefined.

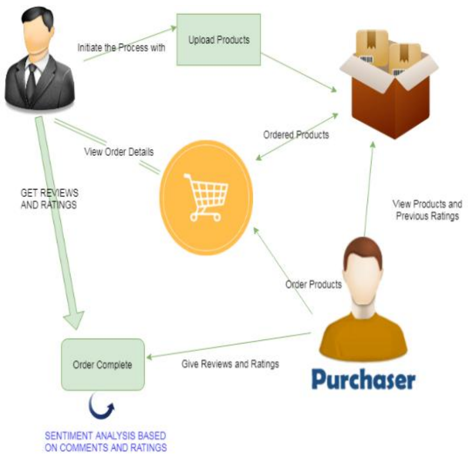
Weak Supervision

This module provides the convenience to admin for supervision of the ratings and reviews. It supervises the given rating is high for positive comment or low ratings for negative comments. It shows the admin that how user rated for the products. It shows the comments and rating on the products.

Graphical Analysis

In this phase of the Implementation user can get the clear picture analysis of the products ratings and reviews. Various factors take into consideration for the graph analysis. In this phase plot the charts like pie graph, bar chart and so others.

ARCHITECTURE



ALGORITHM

In machine learning, naive Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong (naive) independence assumptions between the features. Naive Bayes has been studied extensively since the 1950s. It was introduced under a different name into the text retrieval community in the early 1960s,[1]:488 and remains a popular (baseline) method for text categorization, the problem of judging documents as belonging to one category or the other (such as spam or legitimate, sports or politics, etc.) with word frequencies as the features. With appropriate pre-processing, it is competitive in this domain with more advanced methods including support vector machines. It also finds application in automatic medical diagnosis. Naive Bayes classifiers are highly scalable, requiring a number of parameters linear in the number of variables (features/predictors) in a learning problem. Maximum-likelihood training can be done by evaluating a closed-form expression, which takes linear time, rather than by expensive iterative approximation as used for many other types of classifiers. In the statistics and computer science literature, naive Bayes models are known under a variety of names, including simple Bayes and independence Bayes. All these names reference the use of Bayes' theorem in the classifier's decision rule, but naive Bayes is not (necessarily) a Bayesian method Naive Bayes is a simple technique for constructing classifiers: models that assign

ISSN: 2278-4632
Vol-11 Issue-01 2021

class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. It is not a single algorithm for training such classifiers, but a family of algorithms based on a common principle: all naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable. For example, a fruit may be considered to be an apple if it is red, round, and about 10 cm in diameter. A naive Bayes classifier considers each of these features to contribute independently to the probability that this fruit is an apple, regardless of any possible correlations between the color, roundness, and diameter features. For some types of probability models, naive Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood; in other words, one can work with the naive Bayes model without accepting Bayesian probability or using any Bayesian methods. Despite their naive design and apparently oversimplified assumptions, naive Bayes classifiers have worked quite well in many complex real-world situations. In 2004, an analysis of the Bayesian classification problem showed that there are sound theoretical reasons for the apparently implausible efficacy of naive Bayes classifiers.Still, a comprehensive comparison with other classification algorithms in 2006 showed that Bayes classification is outperformed by other approaches, such as boosted trees or random forests**.**

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] Y. Bengio. Learning deep architectures for ai. Foundations and trendsR in Machine Learning, 2(1):1–127, 2009.

[2] Y. Bengio, A. Courville, and P. Vincent. Representation learning: A review and new perspectives. IEEE TPAMI, 35(8):1798–1828, 2013.

[3] C. M. Bishop. Pattern recognition and machine learning. springer, 2006.

[4] L. Chen, J. Martineau, D. Cheng, and A. Sheth. Clustering for simultaneous extraction of aspects and features from reviews. In NAACL-HLT, pages 789–799, 2016.

[5] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa. Natural language processing (almost) from scratch. JMLR, 12:2493–2537, 2011.

[6] K. Dave, S. Lawrence, and D. M. Pennock. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In WWW, pages 519–528, 2003.

[7] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman. Indexing by latent semantic analysis. Journal of the American society for information science, 41(6):391, 1990.

[8] X. Ding, B. Liu, and P. S. Yu. A holistic lexicon-based approach to opinion mining. In WSDM, pages 231–240, 2008.

[9] L. Dong, F. Wei, C. Tan, D. Tang, M. Zhou, and K. Xu. Adaptive recursive neural network for target-dependent twitter sentiment classification. In ACL, pages 49–54, 2014.

[10] J. Duchi, E. Hazan, and Y. Singer. Adaptive subgradient methods for online learning and stochastic optimization. JMLR, 12:2121–2159, 2011.

[11] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin. Liblinear: A library for large linear classification. JMLR, 9:1871–1874, 2008.

[12] R. Feldman. Techniques and applications for sentiment analysis. Communications of the ACM, 56(4):82–89, 2013.

[13] J. L. Fleiss. Measuring nominal scale agreement among many raters. Psychological bulletin, 76(5):378, 1971.

[14] X. Glorot, A. Bordes, and Y. Bengio. Domain adaptation for large-scale sentiment classification: A deep learning approach. In ICML, pages 513– 520, 2011.

[15] A. Graves and J. Schmidhuber. Framewise phoneme classification with bidirectional lstm and other neural network architectures. Neural Networks, 18(5):602–610, 2005.

[16] K. Greff, R. K. Srivastava, J. Koutn´ık, B. R. Steunebrink, and J. Schmidhuber. Lstm: A search space odyssey. arXiv preprint arXiv:1503.04069, 2015.

[17] H. Halpin, V. Robu, and H. Shepherd. The complex dynamics of collaborative tagging. In WWW, pages 211–220, 2007.

[18] S. Hochreiter and J. Schmidhuber. Long short-term memory. Neural computation, 9(8):1735–1780, 1997.

[19] M. Hu and B. Liu. Mining and summarizing customer reviews. In SIGKDD, pages 168–177, 2004.

[20] N. Kalchbrenner, E. Grefenstette, and P. Blunsom. A convolutional neural network for modelling sentences. In ACL, 2014.