Neural Networks 58 (2014) 60–67

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

2014 Special Issue

Incorporating conditional random fields and active learning to improve sentiment identification

Kunpeng Zhang[a,1], Yusheng Xie[b,*,1], Yi Yang[b], Aaron Sun[c], Hengchang Liu[d], Alok Choudhary[b]

[a] University of Illinois at Chicago, Chicago, IL, USA
[b] Northwestern University, Evanston, IL, USA
[c] Cloud Research Lab, Samsung Research America, San Jose, CA, USA
[d] University of Science and Technology of China, Hefei, China

ARTICLE INFO

Article history:
Available online 10 May 2014

Keywords:
Conditional random fields
Active learning
Customer reviews
Sentiment analysis

ABSTRACT

Many machine learning, statistical, and computational linguistic methods have been developed to identify sentiment of sentences in documents, yielding promising results. However, most of state-of-the-art methods focus on individual sentences and ignore the impact of context on the meaning of a sentence. In this paper, we propose a method based on conditional random fields to incorporate sentence structure and context information in addition to syntactic information for improving sentiment identification. We also investigate how human interaction affects the accuracy of sentiment labeling using limited training data. We propose and evaluate two different active learning strategies for labeling sentiment data. Our experiments with the proposed approach demonstrate a 5%–15% improvement in accuracy on Amazon customer reviews compared to existing supervised learning and rule-based methods.

1. Introduction

The rapid proliferation of Internet connectivity has led to increasingly large volumes of electronic commerce, resulting in a huge amount of social media data in various forms such as online customer reviews, blog articles, social network comments, and microblog messages (e.g., tweets in Twitter). Analyzing and mining useful information from these data using computational techniques, including social network analysis and web information retrieval, has become an important task. With the current trend, an increasing number of people expresses their opinions publicly via social media platforms. According to two surveys conducted on more than 2000 American adults (Pang & Lee, 2008), we note that:

1. 81% of Internet users (60% of American users) have researched a consumer product online at least once;
2. Among readers of online reviews of restaurants, hotels, and other services (including travel agencies or doctors), between 73% and 87% participants report that previous reviews had a significant influence on their decision to purchase;
3. 32% have provided a rating on a product or service via an online ratings system; and
4. 30% have posted an online comment or review regarding a product or service, including 18% of online senior citizens.

Understanding the sentiment of sentences allows us to summarize online opinions which could help people make informed decisions. Automated sentiment identification has seen great research efforts for many years and has achieved some promising results. On one hand, different machine learning techniques, statistical learning methods, and computational linguistic methods have been developed to recognize sentiments (Hu & Liu, 2004; Xie et al., 2013). On the other hand, some researchers have also proposed rule-based (and unsupervised) methods to improve sentiment classification (Cambria, Schuller, Xia, & Havasi, 2013; Hu, Tang, Gao, & Liu, 2013). All of the state-of-the-art algorithms perform well on individual sentences without considering any context information, but their accuracy is dramatically lower on the document level because they fail to consider context. New algorithms are needed to analyze sentiment in longer documents.

* Corresponding author. Tel.: +1 7739515319.
E-mail addresses: kzhang6@uic.edu (K. Zhang), yushengxie2011@u.northwestern.edu, yxi389@eecs.northwestern.edu (Y. Xie), yya518@eecs.northwestern.edu (Y. Yang), a.sun@samsung.com (A. Sun), hcliu@ustc.edu.cn (H. Liu), choudhar@eecs.northwestern.edu (A. Choudhary).
[1] Authors contributed equally.

http://dx.doi.org/10.1016/j.neunet.2014.04.005
0893-6080/

There are many difficulties owing to the special characteristics and diversity in sentence structure, in which people express their opinions. For example, one sentence may express multiple sentiments though the speaker may emphasize one part, as in "*The color of this camera is pretty good, but it is too expensive comparing to similar products from other manufactures*". Also, sarcastic sentences express opinions differently from what texts would suggest in literal, and many sentences express their author's opinions indirectly through comparison. For example, the sentence "*In terms of customer service, Nikon wins over Canon, hands down.*" expresses the reviewer's preference over Nikon cameras, which can be positive or negative depending on whether Nikon or Canon is the main subject of the document to which the sentence belongs. Such sentences explicitly show positive or negative sentiments, but their implicit sentiments are different if they are placed into a particular context. Some typical representatives are sarcastic sentences. Capturing relationships among such sentences in a document is therefore a particular challenge.

In addition, complicated sentence structure and Internet slang make sentiment analysis even more challenging. In this paper, we not only consider syntax that may influence the sentiment, including newly emerged Internet language, emoticons, positive words, negative words, and negation words, but we also incorporate information about sentence structure, like conjunction words and comparisons. The context around a sentence plays an important role in determining the sentiment; e.g., a compound sentence is more likely to be positive if both sentences before and after are positive. Therefore, we employ a conditional random fields (CRF) method to capture syntactic, structural, and contextual features of sentences. Our experimental results on Amazon customer reviews and Facebook comments show improved accuracy compared to supervised and rule-based methods.

Furthermore, labeling sentiments manually is expensive. Often a large number of labels are necessary when training a probabilistic sentiment model with realistic complexity. Therefore, we apply active learning to tag sequences of unlabeled sentences that are most informationally valuable to the model. We propose two different strategies to select "label-worthy" data with high uncertainty for human beings to label, and our experimental results on customer reviews demonstrate faster convergence compared to baselines. This active learning strategy is especially useful when human effort, compared to data availability (i.e., big data), becomes a scarce resource.

2. Literature

Recently, there has been a lot of research on sentiment analysis using techniques ranging from rule-based, bag-of-words approaches to machine learning techniques. The analyzed subjects range from long documents to short sentences.

2.1. Classifying document sentiment

Document sentiment classification is the analysis/classification of text sentiment on a multi-sentence document (e.g., product reviews or blog articles) as positive or negative (Mcdonald, Hannan, Neylon, Wells, & Reynar, 2007; Pang, Lee, & Vaithyanathan, 2002; Turney, 2002). In Choi and Cardie (2008), the authors present a novel approach based on compositional semantics that incorporates sentence structure into the learning procedure. They also find that "content-word negators" play an important role in determining expression-level polarity. Further in Xie et al. (2013), the authors expand the rules and consider the specialty of social media data to improve sentiment classification. Machine learning has also been widely used to identify sentiments of sentences. In Pang et al. (2002), the authors employ machine learning techniques to classify movie reviews by overall sentiment. Their results show that three machine learning methods (Naïve Bayes, maximum entropy classification, and support vector machine) do not perform as well on sentiment classification as on traditional topic-based categorization. However, an important aspect in document sentiment is the consideration of inter-sentence dynamics, which has not yet been systematically handled in previous works other than several specific rules proposed in Choi and Cardie (2008) and Xie et al. (2013).

2.2. Classifying sentence sentiment

Another important research direction is classifying sentences as positive subjective, negative subjective, or objective (Kim & Hovy, 2004; Riloff & Wiebe, 2003; Wiebe, Bruce, & O'Hara, 1999; Wiebe & Wilson, 2002; Wilson, Wiebe, & Hwa, 2004; Yu & Hatzivassiloglou, 2003). In Narayanan, Liu, and Choudhary (2009), the authors present linguistic analysis of conditional sentences, and build some supervised learning models to determine if sentiments expressed on different topics in a conditional sentence are positive, negative or neutral. The more general problem of rating inference, where one must determine the authors' evaluation with respect to a multi-point scale (e.g., one to five "stars" for a review) can be viewed simply as a multi-class text categorization problem. Predicting degree of positivity provides more fine-grained rating information. At the same time, it is an interesting learning problem in itself.

There have been studies on building sentiment lexicons to define the strength of word sentiment, which is largely the foundation for accurate sentence sentiment. Esuli and Sebastiani (2006) constructed a lexical resource, SentiWordNet, a WordNet-like lexicon emphasizing sentiment orientation of words and providing numerical scores of how objective, positive, and negative these words are. However, lexicon-based methods can be tedious and inefficient and may not be accurate due to the complex cross-referencing in dictionaries like WordNet. The sentiment scoring approach in Liu and Seneff (2009) makes use of collective data such as user star ratings in reviews. By associating user star ratings and frequency with each phrase extracted from review texts, they can easily associate numeric scores with textual sentiment. They propose an approach for extracting *adverb–adjective–noun* phrases based on clause structure obtained by parsing sentences into a hierarchical representation. They also propose a robust general solution for modeling the contribution of adverbials and negation to the score for degree of sentiment.

2.3. Deriving concept-level sentiment

Thanks to the availability of quality lexical resources, modern sentence-level sentiment can be classified very accurately. It is considerably harder to achieve that same with document-level classification because of the ambiguity introduced by twisting context in a document. Well written articles, even with strong opinions, usually lay down evidence for both arguments because defeating counterarguments is often a quite effective way of establishing one's own.

Recognizing the difficulty in summarizing an entire document into a sentiment class, concept-level sentiment analysis (Cambria, Olsher, & Rajagopal, 2014) is recently developed. Concept-level analysis aims to infer the semantics and sentics associated with natural language opinions. It conducts the analysis on fine-grained, feature-based, and possibly comparative entities. For example, we are given a document that reviews and compares an iPhone and a Samsung phone. Concept-level analysis associates sentiment with product features like the smart phone's screen or even sub-features like the color/resolution of the screen.