Information Processing and Management 52 (2016) 36–45

Contents lists available at ScienceDirect

Information Processing and Management

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

Polarity shift detection, elimination and ensemble: A three-stage model for document-level sentiment analysis

Rui Xia[a,d], Feng Xu[b], Jianfei Yu[a], Yong Qi[a,b], Erik Cambria[c,*]

[a] School of Computer Science and Engineering, Nanjing University of Science and Technology, China
[b] School of Economics and Management, Nanjing University of Science and Technology, China
[c] School of Computer Engineering, Nanyang Technological University, Singapore
[d] State Key Laboratory for Novel Software Technology, Nanjing University, China

A R T I C L E I N F O

Article history:
Received 26 June 2014
Revised 31 March 2015
Accepted 10 April 2015
Available online 11 November 2015

Keywords:
Sentiment analysis
Sentiment classification
Polarity shift

A B S T R A C T

The polarity shift problem is a major factor that affects classification performance of machine-learning-based sentiment analysis systems. In this paper, we propose a three-stage cascade model to address the polarity shift problem in the context of document-level sentiment classification. We first split each document into a set of subsentences and build a hybrid model that employs rules and statistical methods to detect explicit and implicit polarity shifts, respectively. Secondly, we propose a polarity shift elimination method, to remove polarity shift in negations. Finally, we train base classifiers on training subsets divided by different types of polarity shifts, and use a weighted combination of the component classifiers for sentiment classification. The results on a range of experiments illustrate that our approach significantly outperforms several alternative methods for polarity shift detection and elimination.

1. Introduction

The volume of user-generated text on the Web in the form of reviews, blogs, and social networks has grown dramatically in recent years. This was mirrored by an increasing interest, from both the academic and the business world, in the field of sentiment analysis, which aims to automatically extract sentiment from natural language text and can be broadly categorized into knowledge-based (Cambria et al., 2013) or statistics-based (Cambria et al., 2013b). While the former leverages on the use of ontologies (Gangemi, Presutti, & Reforgiato, 2014) and semantic networks (Cambria, Hussain, Havasi, & Eckl, 2010a) to infer sentiments from text in an unsupervised way, the latter focuses on machine learning (Poria, Gelbukh, Cambria, Hussain, & Huang, 2014; Xia et al., 2015) and clustering techniques (Cambria, Mazzocco, Hussain, & Eckl, 2011) to detect polarity in a supervised way.

In standard practice, sentiment analysis is considered as a special case of text classification, where a review text is normally represented by a bag-of-words (BOW) model. Then, statistical machine learning algorithms, such as Naïve Bayes, maximum entropy classifier, and support vector machine (SVM) are used for classification. However, the BOW model disrupts word order, breaks the syntactic structures and discards some semantic information of the text. Therefore, it brings about some fundamental deficiencies including the polarity shift problem. Polarity shift refers to a linguistic phenomenon in which the polarity of sentiment can be reversed (i.e., positive to negative or vice versa) by some special linguistic structures

* Corresponding author.
E-mail address: cambria@ntu.edu.sg (E. Cambria).

http://dx.doi.org/10.1016/j.ipm.2015.04.003
0306-4573/

called polarity shifters, e.g., negation ("*I don't like this movie*") and contrast ("*Fairly good, but not my style*"). Obviously, in the BOW model, it is hard to capture the sentiment reversion caused by polarity shifters, because two sentiment-opposite texts (e.g., "*I don't like this movie*" and "*I like this movie*") are regarded to be very similar in the BOW representation.

Several approaches have been proposed in the literature to address the polarity shift problem (Das, 2001; Choi, 2008; Kennedy & Inkpen, 2006; Li, Chen, & Zhou, 2010; Li & Zong, 2009; Na, Khoo, & Zhou, 2004; Polanyi, 2004; Wilson & Hoffmann, 2005). However, most of them focused on either modeling polarity shift in phrase/subsentence-level sentiment classification, or encoding polarity shift in rule-based term-counting methods. Even there were few of them dealing with polarity shift by using machine learning methods for document-level sentiment classification, their performances were not satisfactory, e.g., the improvements were less than 2% after considering polarity shift in (Li et al., 2010).

In this work, we propose a three-stage model, namely Polarity Shift Detection, Elimination and Ensemble (PSDEE), to address polarity shift for document-level sentiment classification. Firstly, we propose a hybrid polarity shift detection approach, which employs a rule-based method to detect some polarity shifts such as explicit negations and contrasts, and a statistical method to detect some implicit polarity shifts such as sentiment inconsistencies. Secondly, we propose a novel polarity shift elimination algorithm to eliminate polarity shifts in negations. For example, the review "*this movie is not interesting*" is reversed to "*this movie is boring*". It can make the BOW representation more feasible due to the elimination of negations. Finally, we separate the training and test data into four component subsets, i.e., negation subset, contrast subset, sentiment-inconsistency set as well as polarity-unshifted subset, and train the base classifiers based on each of the component subset. A weighted ensemble of four component predictions are finally used in testing, with the motivation to distinguish texts with different types of polarity shifts such that the polarity-unshifted part will have a higher weight, while the polarity-shifted part will have a lower weight in sentiment prediction. We systematically evaluate our PSDEE model by conducting experiments on four sentiment datasets, three kinds of classification algorithms and two types of features. The experimental results prove the effectiveness of our PSDEE model across different settings.

The rest of the paper is organized as follows: Section 2 presents the motivation; in Section 3, we introduce our PSDEE model in detail by discussing (a) the hybrid polarity shift detection method, (b) the negation elimination approach, and (c) the polarity-shift-based ensemble model; experimental results are reported and analyzed in Section 4; we review related work in Section 5; finally, Section 6 draws the conclusions.

2. Motivation

2.1. How to detect different types of polarity shifts?

Polarity shifters, also called "valence shifter" in (Polanyi & Zaenen, 2004) and "sentiment shifter" in (Liu, 2012) are words and phrases that can change sentiment orientations of texts. Polarity shift is a complex linguistic structure that may include explicit negations, contrasts, intensifiers, diminishers, irrealis, etc. (Taboada, Brooke, Tofiloski, Voll, & Stede, 2011). Li et al. (2010) have conducted a statistic on the distribution of different types of polarity shift, and reported that explicit negations and contrasts covers more than 60% polarity shift structures.

Negation is the most common type of polarity shifts. For example, in the review:

Review 1 (Explicit Negation)**.** "*I don't like this movie*", the negator "*don't*" shifts the polarity of the sentiment word "*like*". It usually has explicit hints (i.e., negators) in negation. Therefore, we can capture the explicit negation by using some rule-based methods based on the presence of some pre-defined negators.

Contrast is another important class of polarity shifts. For example in the review:

Review 2 (Explicit Contrast)**.** *"Fairly good acting, but overall a disappointing movie"*, the contrast indicator "but" shifts the sentiment polarity of the previous phrase "Fairly good acting". Similar as explicit negations, we may also use rule-based method to detect the explicit contrasts according to some pre-defined contrast indicators.

While some polarity shift structures such as explicit negations and contrasts are relative easy to detect, there still exists a large part of implicit polarity shifts that are very hard to detect based on simple rule-based methods. For example in the review:

Review 3 (Sentiment Inconsistency)**.** *"I don't like this movie. Great actor, awful scenario*", the first phrase "*I don't like this movie*" expresses a negative sentiment toward the whole film, the second phrase "*great actor*" shows a positive sentiment toward acting, and the third phrase "*awful scenario*" expresses the negative sentiment toward the aspect of scenario. In this case, people hold an opposite opinion toward one subordinate aspect, which is opposite to the sentiment of the whole review. We call this type of polarity shift "sentiment inconsistency". Pang, Lee, and Vaithyanathan (2002) referred to this problem as "thwarted expectation", which is synonymous to inconsistent or mixed sentiment patterns in the review text. This phenomenon is especially common in long review texts, where people might have different opinions toward different aspects of one product. But in sentiment inconsistencies, the opinion is inconsistent to that in its neighbors, and is always contrary to the sentiment expressed on the product overall. In this case, there are not explicit hints for polarity shift detection. Nevertheless, we could use a statistical method to detect them.