5D-7

Sentiment Classification by Capturing User Preferences across Targets

Wenliang Gao‡ Nobuhiro Kaji† Naoki Yoshinaga† Masaru Kitsuregawa†
†Institute of Industrial Science, University of Tokyo
‡Graduate School of Information Science and Technology, University of Tokyo

1 Introduction

Behind the increase of subjective text on the Web, sentiment classification has received much attention from NLP researchers. This task is defined to classify a subjective text written by a reviewer towards a certain target into his/her sentiment. Although a substantial effort has been devoted to this task, the researchers have usually relied on textual clues extracted from the review content to solve this task. In other words, they assume that the review has been written by anonymous person towards an anonymous target.

This formalization, however, does not reflect the true picture of actual sentiment classification scenario, and naturally raises a question on whether the information on the reviewer or the target product of the review could contribute to the accurate classification. Concretely speaking, given a review written by a reviewer towards a target, if we know, for example, that the user has positive/negative sentiment towards some other target, could it help us predicting his/her sentiment towards the target product in question?

We human can in fact guess a possible sentiment of a user towards an unseen target by observing the other target s/he dis/likes. Imagine a man who likes “iPod shuffle” and “iPad mini” writes a review for “iPod nano”. We expect that he would also favour “iPod nano” since it shares several properties (such as manufacturer and design philosophy) with “iPod shuffle” or “iPad mini”. Similarly, a man who likes FC Barcelona would dislike RealMadrid C.F, since they are rivals. If we are aware of the existence of the users and could capture this sort of correlation between user preferences towards different targets, it should provide a good prior to predict the his/her sentiment.

This paper is motivated from the above observation and proposes a method of capturing correlation between user preferences toward different targets in sentiment classification. Our method assumes that an input review is accompanied with the reviewer and the sentiment target, and induces a feature that captures the preference correlation between targets. For a given document, a new feature is activated when we observe that s/he gives sentiment for some other target. For example, if the user likes “iPad” is oberved, given a questioning product “iPod” the feature expressing this information is set to 1. We call this feature user preference feature. If the training data includes reviews for a pair of products given by the same user and their sentiment has statistically biased, it could help the classification.

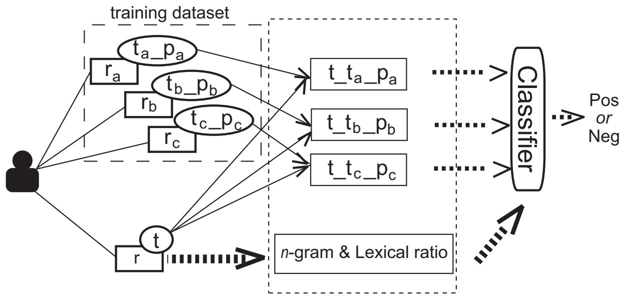

Figure 1: Pipeline for proposed method

2 Proposed Method

We follow a standard supervised approach [1, 3] to sentiment classification and incorporate a new feature that captures a correlation between user preferences towards different targets. The output is a class which can be “negative” or “positive” that expresses sentiment.

The reviews are separated into training and testing partitions. The training set is used for learning parameters and also considered as prior knowledge from which the preference correlation is learned. The testing set is considered unseen data, then it is only used for testing.

The pipeline is shown in figure 1. As illustrated in this figure, for any given document we extract the n-gram features and lexical ratio features in the same manner as previous studies [3, 1]. We extract proposed features by considering other reviews written by the reviewer in the training set. Each such review r is represented as a triple (c_r, u_r, t_r), where c_r is the content, u_r is the user and t_r is the target of the document. The annotation of sentiment polarity for r is denoted as $p_r \in \{pos, neg\}$. With the prior knowledge we construct new features $t_r\_t_i\_p_i$, This feature activate when when u_r write a review towards other targets t_i with sentiment p_i in the training set.

We use n-gram and lexical ratio features as baseline. The proposed method is by adding newly constructed preference features.

2.1 baseline features

For n-gram features, we use indicators of unigrams and bigrams. A hand-made stop word list is used to eliminate functional words. For lexical ratio feature, we use lexicon proposed by [4]. It is computed as:

$$ratio = \frac{posN - negN}{|posN| + |negN|}$$

where $negN/posN$ is the positive/negative word number in the review content c_r.

Method	Accuracy (%)
Baseline	73.83
Proposed	**75.94**
Speriosu[2]	71.20

Table 1: Accuracy on tweet dataset

2.2 User Preference Features

Given a review $r = (c_r, t_r, u_r)$, user u_r posted other documents containing sentiment can be represented as a tuple set $S_{u_r} = \{((c_i, t_i, u_i), p_i) | u_i = u_r \cap i \neq r\}$, where p_i is the annotated sentiment polarity for review i. At last, for each tuple in S_{u_r}, we activate a feature $t_r\_t_i\_p_i$.

For example, given review $r = (c_r, \text{“}iPad\ mini\text{”}, u_r)$, user u_r also posted two other reviews in training dataset naming $r_1 = (c_1, \text{“}SamsungNexus'', u_r)$ with polarity “neg” and $r_2 = (c_2, \text{“}iPod'', u)$ with polarity “pos”. Then we activate two new features as

$$f_1 = iPadMini\_SamsungNexus\_neg$$
$$f_2 = iPadMini\_iPod\_pos.$$

These features capture the relation between user preference for different targets' sentiment and the current target's sentiment based on all the observable users' choices.

3 Experiments

We evaluate the accuracy of the polarity classification. The accuracy is calculated as the number of correctly classified review number divided by the overall number of reviews. We use LIBLINEAR [*] as classifier considering that it learns very fast on massive amount of features. The text segmentation and tokenization are performed by OpenNLP package [†].

3.1 Dataset

The dataset is tweet dataset collected and annotated by [2]. Each tweet datum includes content, user and target information. It is about a debate on “HCR” (Health Care Reform). 10 political entities are mentioned in this dataset as sentiment targets. These include the Health Care Reform policy (“HCR”), Democratic Party (“Dem”), Conservatives (”Conservative”), The President Barack Obama (“Obama”) and Republican Party (“GOP”).

3.2 Result

The results comparison is shown in Table 1. We use baseline features and supervised classifier as baseline. The proposed method is by adding user preference feature to the baseline methos. [2]'s result is also listed as “Speriosu.” Hyper-parameter for classifier is tuned using development datasets. Proposed method outperforms “Speriosu” and baseline by over 4 and 2 percentage respectively.

*http://www.csie.ntu.edu.tw/ cjlin/liblinear/
†http://opennlp.apache.org/

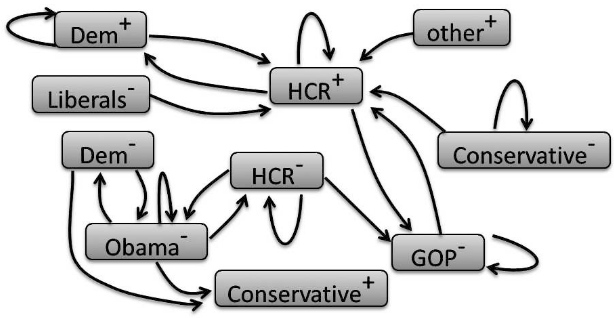

Figure 2: Correlation graph constructed by 40 most weighted preference features. “+”/“-” means pos/neg polarity

3.3 Result Analysis

We further analyze the weights of features learned by LIBLINEAR. The new feature $t_1\_t_2\_p_2$ can be expressed as a linkage of two nodes $t_1\_p_1$ and $t_2\_p_2$ in a correlation graph. The weight for this linkage and the p_1 value is decided by LIBLINEAR learned weight $w^c_{t_1\_t_2\_p_2}$ for feature $t_1\_t_2\_p_2$. Weight is positive means that the feature contributes to classifying a tweet, which contains sentiment towards the target t_1, to “positive,” and vice versa. Then we can translate the weight $w^c_{t_1\_t_2\_p_2}$ into the relatedness of the tuple (t_2, p_2) to tuple (t_1, p_1), where p_1 and link weight $w^l_{(t_1,p_1),(t_2,p_2)}$ are computed as:

$$p_1 = \begin{cases} pos, \ if\ w_i > 0 \\ neg, \ if\ w_i <= 0 \end{cases}, \ w^l_{(t_1,p_1),(t_2,p_2)} \propto |w^c_{t_1\_t_2\_p_2}|$$

The graph is shown in Figure 2. We listed the 40 highest weights learned by the classifier. The relatedness correctly reflect the real scenario. Such as people who support the “Democratic” tend to vote for “HCR”.

References

[1] P. Bo, L. Lee, and V. Shivakumar. Thumbs up? sentiment classification using machine learning techniques. In *EMNLP*, pages 79–86, 2002.

[2] M. Speriosu, N. Sudan, S. Upadhyay, and J. Baldridge. Twitter polarity classification with label propagation over lexical links and the follower graph. In *EMNLP workshop on Unsupervised Learning in NLP*, pages 53–63, 2011.

[3] P. D. Turney. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *ACL*, pages 417–424, 2002.

[4] T. Wilson, P. Hoffmann, S Swapna, J Kessler, J Wiebe, Y. Choi, C. Cardie, E. Rillof, and S. Patwardhan. Opinionfinder: A system for subjectivity analysis. In *EMNLP*, pages 34–35, 2005.