Co-training and Visualizing Sentiment Evolvement for Tweet Events

Shenghua Liu
Institute of Computing Technology, CAS
Beijing, China 100190
liushenghua@ict.ac.cn

Wenjun Zhu, Ning Xu
Wuhan University of Technology, Wuhan, Hubei, China 430070
zwenjun@126.com, xuning@whut.edu.cn

Fangtao Li
Google Inc.
Beijing, China 100084
fangtao06@gmail.com

Xue-qi Cheng, Yue Liu, Yuanzhuo Wang
Institute of Computing Technology, CAS
Beijing, China 100190
{cxq, liuyue, wangyuanzhuo}@ict.ac.cn

ABSTRACT
Sentiment classification on tweet events attracts more interest in recent years. The large tweet stream stops people reading the whole classified list to understand the insights. We employ the co-training framework in the proposed algorithm. Features are split into text view features and non-text view features. Two Random Forest (RF) classifiers are trained with the common labeled data on the two views of features separately. Then for each specific event, they collaboratively and periodically train together to boost the classification performance. At last, we propose a "river" graph to visualize the intensity and evolvement of sentiment on an event, which demonstrates the intensity by both color gradient and opinion labels, and the ups and downs of confronting opinions by the river flow. Comparing with the well-known sentiment classifiers, our algorithm achieves consistent increases in accuracy on the tweet events from TREC 2011 Microblogging and our database. The visualization helps people recognize turning and bursting patterns, and predict sentiment trend in an intuitive way.

Categories and Subject Descriptors
I.2.7 [Natural Language Processing]: *Text analysis*.

General Terms
Algorithms, Design, Experimentation.

Keywords
co-training, sentiment analysis, visualization, Microblog events

1. INTRODUCTION
The booming Twitter service attracts more people to post their feelings and opinions on some trending topics or events online. Sentiment analysis plays an import role to help people understand that. Recent sentiment analysis studies show many interests in large-scale tweets or blogs [1-3]. Some studies [2, 3] especially focus on the sentiment evolvement of tweet events.

However, with poorly designed sentiment visualization, it prevents people to grasp the insights, without reading the large classified list of unstructured tweets. The opinion triangle and ring [4] used periodic pattern, which is not applicable to visualize the sentiment evolvement of event series. Alper et al. Hao et al. [5] used pixel cell-based sentiment calendars and high density geo maps for visualization. Nevertheless, those visualizations cannot show the dynamics and trend of sentiment over time series.

This paper is supported by NSF of China (No. 61202213, 61173008, 61232010), NGFR 973 Program of China (No. 2013CB329602), and National Science supported planning (No. 2012BAH39B02).

WWW 2013 Companion, May 13-17, 2013, Rio de Janeiro, Brazil.
ACM 978-1-4503-2038-2/13/05.

In our algorithm, we employ co-training framework [6], and two sentiment classifiers with different view of features are collaboratively and periodically trained on tweet stream. In the visualization phase, we propose a "river" graph to intuitively show the sentiment classification results for a tweet event.

2. CO-TRAINING
Unlike the product reviews usually companied with a scoring mechanism that quantifies the overall sentiment, tweets lack labeled data. It is a labor-intensive task to manually label a large number of tweets, thus we can only annotate a small set of them, and use the semi-supervised method to utilize the unlabeled ones to boost the performance. Meanwhile, since a tweet is extremely short, it is necessary to extract more features. Besides the traditional textual features, we also need to explore the non-textual features.

Based on the above observations, we design a two-view semi-supervised method for sentiment classification on tweets, which employs the co-training framework. We start to train the classifiers C_1 and C_2 them on a common set of labeled tweets L, and two views of features separately. Then for every specific event, $C_1$and C_2 classify the incoming tweets in a time period t_1, and select confident ones to augment the labeled set L. And we select the p positive tweets and n negative ones, when the classifiers agree most. Several iterations of co-training are executed, and output final classification results by multiplying the scores from both classifiers. In the next period, with last trained classifiers C_1 and C_2, we continue the co-training iteratively and classifying next tweets in stage t_2. Finally, we obtain the weights $p(t)$, $m(t)$, and $n(t)$ for each incoming tweet t, which denote the probabilities that tweet t belongs to the positive, neutral and negative classes.

Features are split into two views, i.e. textual feature and non-textual feature. The textual feature, PMI-IR [7] for each sentiment word w, is computed as:

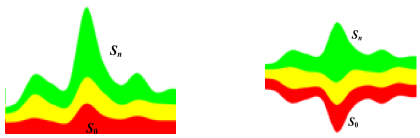
$$\text{PMI-IR}(w) = log_2\left[\frac{hits(w\ NEAR\ "excellent")hits(\text{"poor"})}{hits(w\ NEAR\ "poor")hits(\text{"excellent"})}\right],w \in P$$

where $hits(\bullet)$ is the number of the query results, and P denotes the sentiment dictionary. Here, we use WordNet Affect [8] for sentiment words. Non-textual features include emoticons, temporal features, and punctuation. A set of emoticons from Wikipedia are collected as a dictionary, such as :-), :), (>_<), >:[, :-(, :(, etc. People tend to act differently in the morning and the noon, the beginning and ending of a week or month, spring and winter, etc. Thus we classify the post time into different hours, dates, day of week and months as temporal features. Punctuation marks such as exclamation mark (!), question mark (?), express the emotional intensity. Thus the term frequency of each punctuation mark in a tweet is counted for the feature.

3. SENTIMENT VISUALIZATION

There are some important ingredients that determine the visualization graph, such as graph geometry, layer ordering and coloring, and sentiment words labels.

We define density function ρ_i as the distribution of the number of tweets belonging to sentiment class i. Let the bottom curve function of the visualization graph be S_0. Thus, the upper boundary of the layer of the sentiment class i is $S_i = S_0+\sum_{j=1}^{i}\rho_j$.



(a) (b)

Figure 2. Visualization with ground-based (a) and symmetric (b) graph geometry

By setting function S_0=0 and S_0=–S_n, we can obtain the ground-based and symmetric geometry as Figure 2 shows. Both graphs depict the activeness of discussion via the height or width. However, ground-based graph cannot easily compare the strong and weak between positive and negative sentiments, since their heights are stacked together on the zero line. However, the symmetric graph can grasp the difference through a virtual middle curve through the yellow ribbon. Thus we adopt the symmetric layout around the neutral sentiment in the middle, and mathematically the bottom curve function is derived as

$$S_0=-\frac{1}{2}\sum_{j=1}^{n}\rho_j.$$

Traditionally, there are three sentiment classes. However, such coarse three classes cannot visualize the intensity of people's opinions. Thus, we use fine-grained segments between 0 and 1, and color the layers according to the mapping function as follows.

$$RGB(t) = \begin{cases} \left(\left(1-n(t)\right)\cdot 255, 255, 0\right), & p(t) \geq n(t) \\ \left(255, \left(1-p(t)\right)\cdot 255, 0\right), & p(t) < n(t) \end{cases}$$

In the thoroughly neutral class that $p(t)$ equals to $n(t)$, we simply set the $m(t)$ = 1, and $p(t)$ = $n(t)$ = 0. Thus the middle layer color is yellow with a RGB triple (255, 255, 0).

To extract the sentiment labels, we consider the frequency $f(w)$ and intensity $PMI\text{-}IR(w)$ of an opinion word w, and the font size $F(w)$ of a selected word w is determined as follows, with α as the scaling factor.

$$F(w) = \alpha\bullet PMI\text{-}IR(w)\bullet f(w)$$

4. EXPERIMENTS AND CONCLUSION

We select four events for evaluation shown in Table 1, from our database and TREC 2011 Microblogging. With initially labeled data L_0, we train general RF classifiers on two views for co-training. The accuracies of SVM, standard RF model and co-training algorithm are given for comparison. It is seen that our algorithm outperforms both well-known models, SVM and RF, by achieving 3.8% improvement on the largest event set.

The sentiment visualization results are given in Figure 3 on "Obama Election" from June 2008-May 2009. On the whole, this visualization graph demonstrates that people's sentiments about Obama showed different degrees of fluctuation over time, especially when some influential events occurred. For example, Obama defeated John McCain, was officially elected as the 44th President of the United States and delivered his victory speech, on Nov. 5, 2008. And in late January 2009, he delivered the inaugural address. Besides, the labeled opinion words give the semantic summarization and intensity of sentiment. In the previous work, Thelwall et al [2] used plain curves of volume and sentiment strength to show the sentiment change. Nguyen et al [3] used ratios in a rectangle box to show the dynamics of each sentiment polarity on different objects. Compared to [3] visualizing ratios in a rectangle box and [2], our river graph could show rich information that people care, such as the dynamics of volumes, opinion words, and the sentiment strength distribution along the vertically changing colors from warm to cold.

Table 1. Classifier accuracy on different topics

Topics	Size	L_0	SVM%	RF%	co-train%
US Unemployment	559	135	76.3	78.9	80.2 (+1.6%)
American Train Service	495	189	80.4	79.3	81.6 (+2.9%)
BBC World Service Staff Cuts	729	266	77.8	80.6	83.2 (+3.2%)
Obama Election	41,096	2250	78.1	78.1	81.1 (+3.8%)

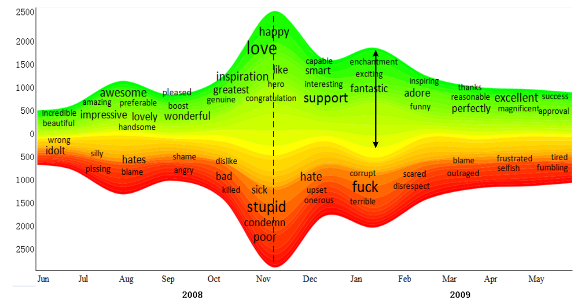


Figure 3. Visualization of about "Obama Election"

In the paper, we propose a periodically co-training algorithm and a "river" graph for classifying and visualizing the sentiment evolvement. Experiments show that our analysis method gains better results in both classification and visualization.

5. REFERENCES

[1] Godbole, N., Srinivasaiah, M., Skiena, S. 2007. Large Scale Sentiment Analysis for News and Blogs. In Proc. of CWSM'2007.

[2] Thelwall, M., Buckley, K. and Paltoglou, G. 2011. Sentiment in Twitter events. Journal of the American Society for Information Science and Technology. Volume 62, Issue 2. pp. 406–418

[3] Nguyen, L. T, Wu, P., Chan, W., Peng, W., Zhang, Y. 2012. Predicting Collective Sentiment Dynamics from Time-series Social Media. In Proc. of WISDOM. p.6.

[4] Wu, Y., Wei, F., Liu, S., Au, N., Cui, W., Zhou, H. and Qu, H. 2010. OpinionSeer: Interactive Visualization of Hotel Customer Feedback, IEEE Trans. on VCG, Vol. 16, No. 6, pp. 1109-1118.

[5] Hao, M., Rohrdantz, C., Janetzko, H., Dayal, U. 2011. Visual Sentiment Analysis on Twitter Data Streams. IEEE Symposium on Visual Analytics Science and Technology, pages 275-276.

[6] Blum, A. and Mitchell, T. 1998.Combining labeled and unlabeled data with co-training. In Proc. of COLT. pp. 92-100.

[7] Turney, P. D. 2002. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In Proc. of ACL. pp. 417-424

[8] Strapparava, C. and Valitutti, A. 2004. Wordnet-affect: an affective extension of wordnet. In Proc. of LREC. pp. 1083-1086.