

Sentiment Visualization on Tweet Stream

Hua Jin

College of Information Science & Technology, Agricultural University of Hebei, China

Email: jinhua923@163.com

Yatao Zhu^{1,2}, Zhiqiang Jin¹, Sandhya Arora³

1 College of Information Science & Technology, Agricultural University of Hebei, China

2 Institute of Computing Technology, Chinese Academy of Sciences, China

3 Meghnad Saha Institute of Technology, Kolkata, India

Email: {yatao1106, sandhyabhagat} @gmail.com

Abstract—Sentiment visualization on tweet topics has recently gained attentions due to its ability to efficiently analyze and understand the people's feelings for individuals and companies. In this paper, we propose a chart, *SentimentRiver*, which effectively demonstrates the dynamics of sentiment involvement on a topic of tweets. The gradient colors of the river flow indicate the variation of topical sentiments, via introducing the membership weight to a sentiment class in a fuzzy mathematical view. Besides, with the value of the point-wise mutual information and information retrieval (PMI-IR), representative sentiment words are extracted and labeled in each time slot of the river flow. In the experiments, we compare *SentimentRiver* on the topic of Obama election, with other statistic charts, which demonstrates its effectiveness for visualizing and analyzing the topical sentiments on tweet stream.

Index Terms—Sentiment visualization, PMI-IR, WEKA, *SentimentRiver*

I. INTRODUCTION

With the rapid development of Internet technology and socialization, people are increasingly accustomed to express their feelings and emotions online. Therefore, emotional information has been aggressively distributed in a variety of social medias, such as product reviews, news comments, microblogs, social networks, etc. However, facing the massive emotional data, people cannot get any overall impression without sentiment extracting and analyzing. Sentiment extraction and analysis in this type of content not only give an emotional snapshot of the online world but also have potential commercial and sociological values for individuals, merchants and even the governments.

Visualization as one of the most efficient sentiment analysis measures provides an intuitive way to exam and analyze the results of auto sentiment classification, which is no longer a passive process that produces images from a set of numbers. In the paper, we design and propose our own flow chart, named *SentimentRiver*, to show the topical sentimental variation over time across a collection of

dynamic tweet stream. *SentimentRiver* is built on the three weights that a tweet belongs to positive, neutral, and negative opinions, which reflects the membership of a tweet belonging to each class. As fuzzy mathematical model shows, each neighboring classes does not clearly bounded by a threshold in reality. Thus a mapping function of the color gradient with the weights is proposed to give a visually demonstration for the fuzzy membership. Random forest [1-2] is selected as the membership function to estimate the weights, learning from the features of the Point-wise Mutual Information and Information Retrieval (PMI-IR), emoticon, post time, etc. Furthermore, the representative sentiment words in each time slot is extracted by the PMI-IR values, and labelled on the *SentimentRiver*.

The rest of the paper is organized as follows. In Section 2, we describe prior works on sentiment analysis in addition to some visualization works. The details of estimating the membership weights and building *SentimentRiver* graph are describe in Section 3. And in Section 4, we describe the experimental results. Finally, conclusions and future work are demonstrated.

II. RELATED WORK

In this section we briefly present some of the research literature related to sentiment analysis and visualization.

Sentiment analysis is a hot topic in the area of Natural Language Processing and text mining in recent years. There are a large amount of works in sentiment classification, most of which focused on handling product or service reviews, and information seeking [1,3-6]. Turney [1] presented an effective unsupervised learning algorithm, called semantic orientation, for classifying reviews as recommended or not recommended. A web-kernel based measurement was proposed as PMI-IR, which is independent to the corpus collection in hand. An opinion-oriented information-seeking system was introduced and gave a relative comprehensive survey of opinion mining and sentiment analysis technologies around the system. Hu and Liu [3] focused on mining opinion features from product reviews. Li et al. [6] predicted the review rating by considering the reviewers and products.

Manuscript received January 16, 2014; revised February 28, 2014

Visualization is becoming an important way to gain insight on the themes, sentiments, and dynamics of complex data. Wu et al [7] proposed the opinion triangle and ring to visualize the hotel reviews of different places and time periods. Alper et al.[8] visualized the overall opinions on product features with the help of OpinionBocks. Nevertheless, those visualization approaches cannot track the evolvement of topical sentiments, since of the dynamics of the topics. Harve et al [9] proposed a prototype system called ThemeRiverTM, which visualized thematic variations over time across a collection of documents. They used colored “currents” flowing within the river represent individual themes. Wattenberg [10] described a new kind of stacked graph, the Streamgraph. This complex layered graph was effective for displaying large data set to a mass audience. A flow chart is proposed to visualize the text and topics of a collection of documents along the time series [11-13]. In the paper, we redesign the flow graph with gradient colors to show the variation of topical sentiments over time across dynamic tweet stream. With the view of fuzzy modeling, the smooth color changing effectively visualizes the membership of a tweet to sentiment classes.

III. SENTIMENT ANALYSIS WITH SENTIMENTRIVER

SentimentRiver is a novel graphical approach which combines a set of visualization techniques with effective sentiment classification approach to help users explore and analysis topical sentiments on large collections of tweets. There are four main ingredients that determine a generalized SentimentRiver chart, and we will explore them in proper order.

A. SentimentRiver Graph Geometry

To describe the geometry precisely, we use the following notation. We model the sentiment series as a set of n real-valued non-negative functions, t_1, \dots, t_n . We define the bottom of the stacked graph as baseline function S_0 . The top of the layer corresponding to the i th sentiment series f_i is therefore given by the function S_i , where

$$S_i = S_0 + \sum_{j=1}^i t_j$$

If we set the baseline function $S_0=0$, the SentimentRiver graph is a traditional stacked graph which based at zero (Figure 1). Considering the goal of our SentimentRiver chart is to visually analyze the tri-polar sentiments (positive, negative, and neutral) in a tweets collection and their changes over time, so it is important for us to judge which one is preponderant between positive sentiment and negative sentiment. But it is difficult to get this information from the traditional stacked graph geometry. Therefore, we adopt a layout symmetric around the neutral sentiment in the middle. It is similar to the ThemeRiver [14-16] layout, which is a pretty symmetric style around x-axis. Mathematically, this can be expressed as: $S_0 + S_n = 0$. With the definition of S_0 ,

$2S_0 + \sum_{i=1}^n t_i = 0$. We get the SentimentRiver resolution for S_0 : $S_0 = -\frac{1}{2} \sum_{i=1}^n t_i$.

Figure 2 presents the SentimentRiver chart with a symmetric layout, which balances the interplay between aesthetics and legibility. In this graph, if the middle “current” has a reclined trend, we know the positive sentiment (top layer) outbalance negative sentiment (bottom layer); otherwise, it means the positive sentiment achieves a dominant position. What’s more, the symmetric layout

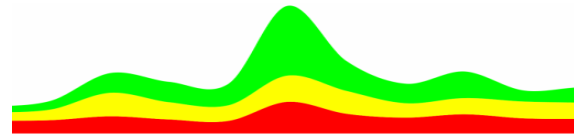


Figure 1. SentimentRiver with traditional stacked graph geometry



Figure 2. SentimentRiver with the symmetric graph geometry

reduced the wiggles between layers and the overall visual distortion. That’s to say, our SentimentRiver chart reduce the wiggles of different layers as much as possible thus present a gradual trend over time, just like the river.

B. Layer Color Gradient

We adopt the RGB color model to present different colors. To form a color with RGB, three colored light beams (one red, one green, and one blue) must be superimposed. Fortunately, our sentiment classification result of each tweet is also determined by three parameters: the positive probability, the negative probability and the neutral probability. For simplicity, we use p , n and m represents these three kinds of probability respectively. And p, n, m satisfy the conditions of $p + n + m = 1.0$. What’s more, we use green, yellow, and red to represent positive, completely neutral and negative respectively, where completely neutral means the tweet is classified as neutral at the probability of 1.0 ($m=1.0$). So the color of tweets t is defined as follows:

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

C. Membership Estimation of Sentiment Classes

To get the membership weights that a tweet belongs to each of the three sentiment classes, we explore the classification models, and select Random Forest as the membership functions [17-18]. We firstly explore some effective features for sentiment classification, then use the supervised learning method on WEKA platform to classi-

fy the tweets to tri-polar sentiments (positive, negative, and neutral).

- Temporal features: we want to track the sentiment evolution trends of one event, so just need to collect tweets about this event within some continuous time. Then, we divide the continuous time into different phases by different level such as one hour, one day, one week or one month, and each time phase represents a different temporal feature value. That is to say, all the tweets in one time phase have the same temporal feature value.
- Semantic-oriented feature: We take advantage of the Point-wise Mutual Information and Information Retrieval algorithm to extract one classification feature, which is called PMI-IR value of a tweet. Considering that the maximum length of a Twitter message is only 140 characters, instead of extracting phrases containing adjectives or adverbs like Turney, we adopt a different method to choose words that need to calculate their PMI-IR values. The method is as follows.

$$PMI - IR(word)$$

$$= \log_2 \left[\frac{hits(wordNEAR"excellent")hits("poor")}{hits(wordNEAR"poor")hits("excellent")} \right]$$

The hits of a word are estimated by issuing queries to AltaVista search engine and noting the number of matching documents. The reference words “poor” and “excellent” are choose from the five star review rating system. And the PMI-IR feature value of a tweet is the average PMI-IR value of all words corresponding to this tweet in set P. In particular, if some tweets have no word in set P, their PMI-IR feature values are set to 0.

In addition to the above features, there are some common features such as conjunction words, negation words, punctuations and unigrams. So we can consider the combination of different features as sentiment classification feature set in later experiments.

D. Sentiment Words Extraction and Labeling

Furthermore, in order to distinguish different layers effectively, we should give some labels on them according to the sentiments they represent, and should pay attention that the labels should not overlap the boundary of layers. So the labels are placed in an optimal spot and added by hand. Particularly, the font sizes of labels are adjusted to fit each layer.

Considering that each layer presents a sentiment, we choose the high frequency sentiment words from all tweets as labels. And the sentiment words are chosen in the process of extracting sentiment features. With regard to the font size of labels, they are determined by the product of their contribution to sentiment classification and their frequency of occurrence. And their contributions to sentiment classification are measured by the absolute value of their PMI-IR. The methods we used to choose the sentiment words and compute the PMI-IR values will be introduced in detail in Section 4. Figure 3

present a labelled SentimentRiver chart of the topic BBC world service staff cuts.

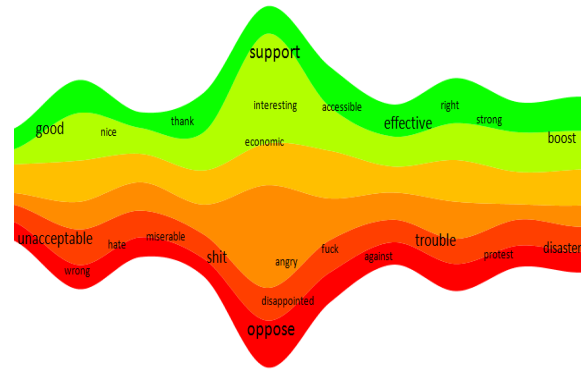


Figure 3. SentimentRiver with labels

IV. EXPERIMENTS

Firstly, we collect millions of tweets via Twitter Streaming API as training data. Then we build our classifiers using different combinations of feature types to observe their individual contributions to the performance. And the classification dataset is about obama, containing 2250 tweets from June 1, 2008 to May 31, 2009. For simplicity, we use NB, SVM, DT and RF on behalf of Naive Bayes, LibSVM, Decision Tree and Random Forest respectively.

In table 1 presents the accuracies achieved by different classifiers trained with different combinations of feature types. When only the temporal features are used, the accuracies are very low. Then with the increase by punctuations features and emoticons features, the accuracies are increased accordingly. And it is obvious from the table that PMI-IR features significantly improve the performance. But when we add the negations features to the feature’s combinations, the accuracies are reduced in NB, DT and RF algorithms. Therefore, we can conclude that

TABLE I.
TYPE SIZES FOR CAMERA-READY PAPERS

| Features | NB | SVM | DT | RF |
|--|------|------|------|------|
| Temporal | 51.6 | 49.2 | 52.1 | 54.3 |
| Temporal +Punctuations | 56.3 | 54.9 | 55.7 | 60.7 |
| Temporal +Punctuations +Emoticons | 61.2 | 61.3 | 61.7 | 66.8 |
| Temporal +Punctuations +Emoticons +PMI-IR | 77.3 | 75.6 | 78.1 | 80.2 |
| Temporal +Punctuations +Emoticons +PMI-IR +Negations | 73.2 | 76.1 | 75.9 | 78.1 |

the best features used for sentiment classification are the combination of temporal, punctuations, emoticons and PMI-IR values.

Next, we use the best feature combinations do experiment on different topics combination with different classifiers. We train our machine learning model using different classification algorithms and test on our data via 10-fold cross-validation. Each time, we use 9 parts as the labelled training data for feature selection and construction of labelled vectors, and the remaining one part is used as a test set. The process was repeated ten times. The classification results are shown in Table 2. Seen form Table 2, Random Forest classifier performs the best. The classification accuracies on all four topics are over 80%. And the other three classifiers do not show obvious differences.

TABLE II.
TYPE SIZES FOR CAMERA-READY PAPERS

| Topics | NB | SVM | DT | RF |
|------------------------|------|------|------|------|
| Obama | 76.5 | 78.9 | 80.5 | 84.3 |
| US Unemployment | 80.3 | 79.3 | 76.3 | 85.6 |
| American Train Service | 77.9 | 80.6 | 72.6 | 83.2 |
| BBC Staff-cuts | 75.2 | 78.1 | 77.6 | 81.1 |

Figure 4 reveals the sentiment changes from June 2008 to May 2009 about the topic of obama. In the SentimentRiver visualization, each layer represents a sentiment of different intensity, which is described by a set of sentiment keywords. These sentiment keywords are distributed along time, summarizing the sentiment evolution over time. The x-axis encodes the time and the y-axis encodes the strength of each sentiment. For each kind of sentiment, the height encodes the number of people that holds this sentiment at a particular time. And from the height of each sentiment and its keywords distributed over time, the user can observe the sentiment evolution over time.

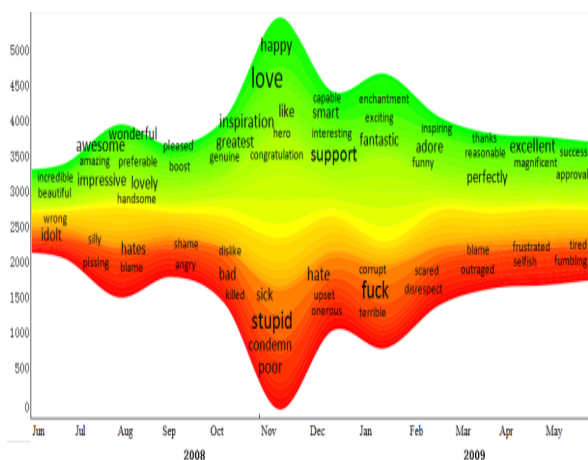


Figure 4. SentimentRiver visualization from June 2008-May 2009 on obama

Figure 4 presents the classification results from the macro-view. We can see some obvious changes in this

graph, such as the increased total river width in early November 2008, which means the number of people that participated in the discussion of Obama reached its peak. Most of this change can be attributed to the significant event that on November 5, 2008, Obama defeated Republican candidate John McCain, was officially elected as the 44th President of the United States and delivered his victory speech.

V. CONCLUSIONS

In this paper, we explored a novel SentimentRiver chart, which combines a set of visualization techniques with effective sentiment classification approach and aims to let users gain useful sentiment information as quickly and as effortlessly as possible, by transforming large collections of tweet sentiment into interactive visualizations. It is designed to progressively disclose increasingly changed sentiment information from topical tweets while continuously providing visual graphical sentiment KEYWORDS. IN FUTURE WORK, WE PLAN TO DEVELOP THE SENTIMENTRiver into a full production system that presents sentiment visualization of different topics for comparison. In addition, we want to do some research work on constructing an unsupervised learning sentiment classifier that applies to any topic.

ACKNOWLEDGMENT

This work is partially supported by Plan Project of Research and Development of Science and Technology of Baoding under Grant No.13ZF098 and No.13ZN025, Youth Foundation of Science and Technology of College of Hebei Province with Grant No.Z2012142, Natural Science Research of Association of Science and Technology of Baoding under Grant No.KX2013A20 and Science and Technology Foundation of Agricultural University of Hebei under Grant No. LG20120604.

REFERENCES

- [1] Turney, P. D. (2001). Mining the Web for synonyms: PMI-IR versus LSA on TOEFL. Proceedings of the 12th European Conference on Machine Learning (pp. 491-502). Berlin: Springer-Verlag.
- [2] Turney, P. D. (2002). Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics ACL'02.
- [3] Hu, M. and Liu, B. (2004). Mining opinion features in customer reviews. In Proceedings of AAAI, pp. 755-760
- [4] Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, vol. 2, nos. 1-2, pp. 1-135, 2008.
- [5] Tang, H., Tan, S. and Cheng. X. (2009). A survey on sentiment detection of reviews. Expert Systems With Applications
- [6] Li, F., Liu, N., Jin, H., Zhao, K., Yang, Q. and Zhu, X. (2011). Incorporating Reviewer and Product Information for Review Rating Prediction. In Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI 2011).
- [7] Wei Xu, Zhi Liu, Tai Wang, Sanya Liu. (2013). Sentiment Recognition of Online Chinese Micro Movie Reviews Us-

ing Multiple Probabilistic Reasoning Model. Journal of Computers. Vol8, No 8, 2013.

- [8] 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, pages 1109-1118.
- [9] Alper, B., Yang, H., Haber, E. and Kandogan, E. (2011) OpinionBlocks: Visualizing Consumer Reviews, IEEE VisWeek 2011 Workshop on Interactive Visual Text Analytics for Decision Making.
- [10] Havre, S., Hetzler, B. and Nowell, L. (2002). ThemeRiverTM: In Search of Trends, Patterns, and Relationships. IEEE Transactions on Visualization and Computer Graphics. 8(1):9-20; 2002.
- [11] Wei, F., Liu, S., Song, Y., Pan, S., Zhou, M. X., Qian, W., Shi, L., Tan, L. and Zhang, Q. (2010). TIARA: A Visual Exploratory Text Analytic System. In Proc. of KDD'10.
- [12] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I. H. (2009). The WEKA data mining software: An update. SIGKDD Explorations, 11(1):10-18, 2009.
- [13] Jianfang Wang, Xiao Jia, Longbo Zhang. (2013). Identifying and Evaluating the Internet Opinion Leader Community Through k-clique Clustering. Journal of Computers. Vol8, No 9, 2013.
- [14] Go, A., Huang, L. and Bhayani, R. (2009). Twitter Sentiment Analysis. CS224N - Final Project Report June 6, 2009.
- [15] Go, A., Bhayani, R. and Huang, L. (2009). Twitter sentiment classification using distant supervision. Technical report, Stanford Digital Library Technologies Project.
- [16] Kouloumpis, E., Wilson, T. and Moore, M. (2011). Twitter Sentiment Analysis: The Good, the Bad and the OMG. Proceedings of the Fifth International AAI Conference on Weblogs and Social Media, 2011.
- [17] Tumasjan, A., Sprenger, T.O., Sandner, P.G. and Welp, I.M. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. In Fourth International AAI Conference on Weblogs and Social Media, Washington, D.C.
- [18] Guoyong Mao, Ning Zhang, Jiang Xie. (2013). A Web-oriented Framework for Graph Simplification and Interactive Visualization. Journal of Computers. Vol8, No 12, 2013.



Hua Jin, Jiangsu Province, China, born in 1980. Computer Science M.Sc., graduated from College of Information Science & Technology, Agricultural University of Hebei. Her research interests include mathematical logic, data mining and agricultural informatization.

She is an assistant professor of College of Information Science & Technology, Agricultural University of Hebei.



Yatao Zhu, Hebei Province, China, born in 1978. A Ph.D. candidate of Institute of Computing Technology, Chinese Academy of Sciences. His research interests include computer architecture, SoC, social computing and agricultural informatization. He is an assistant professor of College of Information Science & Technology, Agricultural University of Hebei.



Zhiqiang Jin, Hebei Province, China, born in 1978. Computer Science M.E., graduated from College of Information Science & Technology, Agricultural University of Hebei. His research interests include data mining and agricultural informatization.

He is an associate professor of Agricultural University of Hebei.

Sandhya Arora, She is currently working as Assistant Professor in Department of Computer Science & Engineering at Meghnad Sah Institute of Technology, Kolkata, WB, India.