Scientific Journal of Impact Factor (SJIF): 4.72
e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 4, Issue 3, March -2017

TASC:Topic-Adaptive Sentiment Classification on Dynamic Tweets

[1]Prof. Dyaneshwar Kudande, [2]Swapnil Desale, [3]Abhishek Kolte

Computer Department, SRTTC, Pune, India

Abstract —*Sentiment classification is a topic-sensitive task, i.e., a classifier trained from one topic will perform worse on another. This is especially a problem for the tweets sentiment analysis. Since the topics in Twitter are very diverse, it is impossible to train a universal classifier for all topics. Moreover, compared to product review, Twitter lacks data labeling and a rating mechanism to acquire sentiment labels. The extremely sparse text of tweets also brings down the performance of a sentiment classifier. In this paper, we propose a semi-supervised topic-adaptive sentiment classification (TASC) model, which starts with a classifier, built on common features and mixed labeled data from various topics. It minimizes the hinge loss to adapt to unlabeled data and features including topic-related sentiment words, authors' sentiments and sentiment connections derived from "@" mentions of tweets, named as topic-adaptive features. Text and non-text features are extracted and naturally split into two views for co-training. The TASC learning algorithm updates topic-adaptive features based on the collaborative selection of unlabeled data, which in turn helps to select more reliable tweets to boost the performance. We also design the adapting model along a timeline (TASC-t) for dynamic tweets. An experiment on 6 topics from published tweet corpuses demonstrates that TASC outperforms other well-known supervised and ensemble classifiers. It also beats those semi-supervised learning methods without feature adaption. Meanwhile, TASC-t can also achieve impressive accuracy and F-score. Finally, with timeline visualization of "river" graph, people can intuitively grasp the ups and downs of sentiments' evolvement, and the intensity by color gradation.*

Keywords--*Sentiment classification, social media, topic-adaptive, cross-domain, multiclass SVM, adaptive feature*

I. INTRODUCTION

THE booming Micro blog service, Twitter, attracts more people to post their feelings and opinions on various topics. The posting of sentiment contents can not only give an emotional snapshot of the online world but also have potential commercial, financial and sociological values. However, facing the massive sentiment tweets, it is hard for people to get overall impression without auto-matic sentiment classification and analysis. Therefore, there are emerging many sentiment classification works showing interests in tweets.Topics discussed in Twitter are more diverse and unpredictable. Sentiment classifiers always dedicate themselves to a specific domain or topic named in the paper. Namely, a classifier trained on sentiment data from one topic often performs poorly on test data from another. One of the main reasons is that words and even language constructs used for expressing sentiments can be quite different on different topics. Taking a comment "read the book" as an example, it could be positive in a book review while negative in a movie review. In social media, a Twitter user may have different opinions on different topics. Thus, topic adaptation is needed for sentiment classification of tweets on emerging and unpredictable topics. Previous works explicitly borrowed a bridge to connect a topic-dependent feature to a known or common feature. Such bridges are built between product reviews by assuming that the parallel sentiment words exist for each pair of topics, such as books, DVDs, electronics and kitchen appliances. However, it is not necessarily applicable to topics in Twitter, especially the unpredictable ones. It is worth mentioning that detecting and tracking topics from tweets is another research topic. Ad-hoc Micro blog search in Text Retrieval Conference (TREC) 2011 [13] and 2012 [14] is hopefully a choice for people to query tweets on emerging topics, and sentiment classification can be conducted afterwards. Unlike product reviews that are usually companied with a scoring mechanism quantifying review sentiments as class labels, there lack labeled data and rating mechanism to generate them in Twitter service. Go et al. used emoticons as noisy labels for Twitter sentiment classification. But the neutral class could not be labeled in this way, and unexpected noise may be introduced only relying on emoticons to label sentiment classes. Semi-supervised approaches have been used for sentiment classification with a small amount of labeled data for other medium than Micro blog. They did not take the special nature of tweets, such as emoticons, users, and networks, to select unlabeled data for training. Tan et al. explored the sentiment correlations affected by users who connect with each other, i.e. social network and @-network formed by users referring to each other with "@" mentions in tweets. However, the work focused on user-level sentiment analysis, while we analyze at the tweet-level.

II. GOALS AND OBJECTIVES

1. To solve the problem of sentiment classification on tweets.
2. To improve the sentiment classification using TASC model.
3. To adapt the sentiment classification dynamics of tweets.
4. Thus, our goal is not to explore the relevance of terms to documents, but to select keywords from the given set of terms to represent the document, such that the quality of answers to triggers/queries is optimized.
5. The focus of these research efforts is on relevance – that is, getting the right set of terms that are most relevant to tweets. In our problem, a set of possibly relevant terms and their relevance to the document are already given by other data processing techniques.

III. EXISTING SYSTEM

The topics in Twitter are very diverse, which makes it impossible to train a universal classifier for all topics. Moreover, compared to product review, Twitter lacks data labeling and rating mechanism to acquire sentiment labels. The extremely sparse text of tweets also brings down the performance of a sentiment classifier. In this paper, we propose a semi-supervised topic-adaptive sentiment classification (TASC) model, which starts with a classifier, built on common features and mixed labeled data from various topics. Acquire sentiment labels.
The main problem of **sentiment analysis on tweets**, because it suffer from the problems of lack of adapting to unpredictable topics and labeled data, and extremely sparse text.

IV. PROPOSED SYSTEM

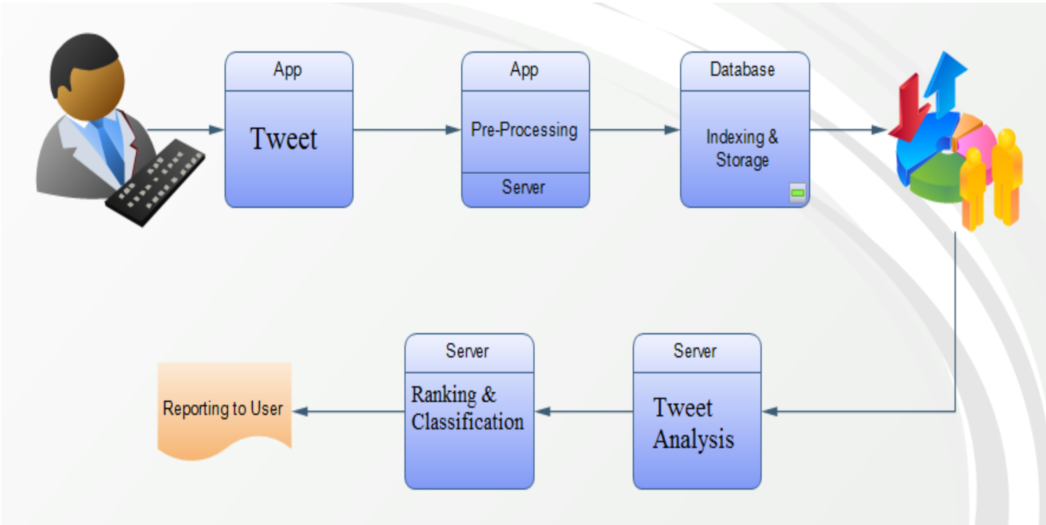

Fig:-System Architecture

Proposed work implements a semi-supervised topic-adaptive sentiment classification (TASC) model, which starts with a classifier built on common features and mixed labeled data from various topics and selected data to adapt the sentiment classifier to the unlabeled data including topic related sentiment words and sentiment connections derived from "@" mentions of tweets, named as topic adaptive features. An enhanced of TASC-t is designed to adapt along a timeline for the dynamics of tweets.

MODULES:-

- TOPIC ADAPTIVE SENTIMENT CLASSIFICATION
- Multiclass SVM

TOPIC ADAPTIVE SENTIMENT CLASSIFICATION

Sentiment classification in the work is treated as a multiclass classification problem, with positive, neutral, and negative expressions. The training data is given in (xi, yi), yi 2 1, _ _ _ ,K, where xi is the feature vector with each element as the value of the corresponding feature, yi is the class that the data belongs to, and k is the number of classes. As we analyze sentiment at tweet document-level, it assumes that a tweet ti belongs to one and only one overall sentiment class, expressing on a single topic from a single opinion holder.
Sometimes opinions are expressed on more than one topics in a post by different holders, so we can divide it into different posts by detecting topic spans and opinion holders.

Multiclass SVM

SVMs model is originally built for binary classification. And there are intuitive ways to solve multiclass with SVMs. The most common technique in practice has been to build K "one-versus-rest" classifiers, and to choose the class which classifies the test data with greatest margin.

V. MOTIVATION AND NEED

- The main problem of **sentiment analysis on tweets**, because it suffer from the problems of lack of adapting to unpredictable topics and labeled data, and extremely sparse text.
- Proposed work implements a **semi-supervised topic-adaptive sentiment classification (TASC) model**, which starts with a classifier built on common features and mixed labeled data from various topics and selected data to adapt the sentiment classifier to the unlabeled data including topic related sentiment words and sentiment connections derived from "@" mentions of tweets, named as topic adaptive features. An enhanced of **TASC-t** is designed to adapt along a timeline for the dynamics of tweets.
- In the GPS navigation system, POI (point of interest) is a geographically anchored pushpin that someone may find useful or interesting, which is usually annotated with texture information (e.g., descriptions and users' reviews).
- Moreover, in many social network services (e.g., Facebook, Flickr), a huge number of geo-tagged photographs are accumulated every day, which can be geo-tagged by users, GPS- enabled smartphones or cameras with a built-in GPS receiver (e.g., Panasonic Lumix DMC-TZ10).
- These uploaded photographs are usually associated with multiple text labels. As a result, in recent years various spatial keyword query models and techniques have emerged such that users can effectively exploit both spatial and textual information of these *spatio textual* objects.

VI. CONCLUSION

Diverse topics are discussed in Microblog services. Sentiment classifications on tweets suffer from the problems of lack of adapting to unpredictable topics and labeled data, and extremely sparse text. Therefore we formally propose an adaptive multiclass SVM model in co-training scheme, i.e., TASC, transferring an initial common sentiment classifier to a topic-adaptive one by adapting to unlabeled data and features. TASC-t is designed to adapt along a timeline for the dynamics of tweets. Compared with the well-known baselines, our algorithm achieves promising increases in mean accuracy on the 6 topics from public tweet corpuses. Besides a well-designed visualization graph is demonstrated in the experiments,
showing its effectiveness of visualizing the sentiment trends and intensities on dynamic tweets.

VII. REFRENCES

[1] S.T. Karamouzis, A. Vrettos, "Sensitivity Analysis of Neural Network Parameters for Identifying the Factors for College Student Success," World Congress on Computer Science and Information Engineering, pp. 671-675, 2009.

[2] J. Han and M. Kamber, "Data Mining: Concepts and Techniques," Morgan Kaufmann, 2000.

[3] T .Chang, Ed. D, "Data Mining: A Magic Technology for College Recruitments," Paper of Overseas Chinese Association for Institutional Research (www.ocair.org), 2008

[4] S. Elayidom, S. Mary Idikkula & J.Alexander, "A Generalized Data mining Framework for Placement Chance Prediction Problems," International Journal of Computer Applications, (0975– 8887) vol.31, no. 3, October 2011.

[5] G.N. Rao, M. Ramachandra, "A Study on the Academic Performance of the Students by Applying K-Means Algorithm," International Journal of Emerging Technologies in Computational and Applied Sciences, vol. 7, no. 4, pp. 365-366, February 2014.

[6] G.N. Rao, S. Nagaraj, "A Study on the Prediction of Student's Performance by applying straight-line regression analysis using the method of least squares," International Journal of Computer Science Engineering, vol. 3, no. 1, January 2014.

[7] Z.J. Kovacic, "Early prediction of student success: Mining student enrollment data," Proceedings of Informing Science & IT Education Conference, 2010. 35 SRTTC, College of engg.2016

[8] M.S.B. PhridviRaj , C.V. GuruRao, "Data mining – past, present and future – a typical survey on data streams," Procedia Technology, vol. 12 , pp. 255 – 263, 2014.

[9] P. Cortez, and A. Silva, "Using Data Mining To Predict Secondary School Student Performance," In EUROSIS, A. Brito and J. Teixeira (Eds.), 2008, pp.5-12.

[10] S. Zhao, E.C.C. Tsang, D. Chen, and X.Wang, "Building a Rule Based Classifier—A Fuzzy-Rough Set Approach,"

