SOUTH ASIAN JOURNAL OF ENGINEERING AND TECHNOLOGY

Full Length Article

EFFICIENT HIGH COVERAGE SOCIAL MEDIA OPINION ANALYSIS USING HASH TAGGER APPROACH

[a]R.Jeevananthem, [a]A.Sriram, [a]R. Sudhakar [a]C.Vijay, [b]R.Sudhakar

[a] *Department of Computer Science and Engineering, Nandha College of Technology, Erode-638052, Tamilnadu, India*
[b] *Assistant Professor, Department of Computer Science and Engineering, Nandha College of Technology, Erode-638052, Tamilnadu, India*

***Corresponding Author**
sudakar.r@nandhatech.org
(R.Sudakar)
Tel.: +9944522778

ABSTRACT: This work proposes a novel Sentiment-Based Enhanced Naïve Bayes to address the information overload problem through information filtering. The proposed framework first applies a Natural Language Processing (NLP) technique to perform sentiment analysis taking advantage of the huge sums of textual data generated in from the social media are predominantly left untouched. Although some current studies do employ review texts, many of them do not consider how sentiments in reviews influence recommendation algorithm for prediction.

General Terms: Your general terms must be any term which can be used for general classification of the submitted material such as Pattern Recognition, Security, Algorithms et. al.

Keywords: Enhanced Naive Bayes, opinion mining and sentiment analysis.

1 INTRODUCTION

Mining the sentiment information in the massive user generated content can help to sense the public's opinions towards a mixture of topics, such as products, brands, disasters, events, celebrities and so going on, and it is useful in many applications. For occurrence, researchers have established that analyzing the sentiments in tweets has the probable to foresee distinction of stock marketplace prices and presidential selection results. Classifying the sentiments of massive micro blog messages is also helpful to substitute or supplement traditional polling, which is expensive and time-consuming.

Product review sentiment analysis can help companies improve their products and services, and help customers make more informed decisions. Analyzing the sentiments of customer generated satisfied is also confirmed useful for client interest removal, personalized recommendation, social publicity, purchaser relation management, and crisis management. As a result, sentiment classification is a hot research topic in both industrial and academic fields.

2 METHODOLOGY

2.1 Opinion mining and Sentiment analysis

With the growing availability and reputation of estimation-rich capital such as online review sites and private blogs, new opportunities and challenges occur as people now can, and do, aggressively use in sequence technologies to search for out and recognize the opinions of others. The unexpected explosion of activity in the area of view mining and sentiment study, which deals with the computational treatment of opinion, sentiment, and partisanship in text, has consequently occurred at least in element as a direct answer to the rush of interest in innovative systems that deal straight with opinions as unparalleled object.

2.2 Opinion Flow

The copy is derived based on two explanations. First, significant users on common media are more expected to transform the opinions of supplementary users. Second, discerning Exposure a elementary theory from media and statement studies, suggests that a user tends to accept an opinion with the purpose of is comparable to his judgment. Thus, incorporate authority and opinion similarity factors into our model.

2.3 Machine Learning Techniques

Using movie reviews as data, notice that standard engine learning techniques definitively do better than human-produced baselines. Nevertheless, the three machine learning methods are engaged (Naive Bays, highest entropy cataloguing, and support vector machines) do not make as well on sentiment classification as on long-established topic-based labeling. To terminate by tentative factors that makes the emotion classification problem more demanding.

Nowadays, very large amounts of information are available in on-line documents. As ingredient of the endeavor to better arrange this information for users, researchers have been actively investigating the problem of regular text cataloguing.

2.4 Domain Revision for Sentiment analysis

sentiment is uttered in a different way in dissimilar domains, and annotating corpora for every probable area of attention is not viable. To look into domain edition for emotion classifiers, focusing on online reviews for dissimilar types of goods.

First, lengthen to response arrangement the recently-proposed structural correspondence learning (SCL) algorithm, dropping the relation fault due to edition between domains by an average of 30% over the original SCL algorithm and 46% over a supervised baseline. Second, to identify a measure of domain comparison that correlates well with the possible for adaptation of a classifier from one area to another.

3 MULTI DOMAIN SENTIMENT CLASSIFICATION

Sentiment classification has been widely known as a highly domain-dependent problem.For example, Pang et al. built sentiment classifiers for movie reviews using machine learning techniques such as SVM and Naive Bayes based on the labeled data of this domain. Lu et al. proposed to construct a domain-specific sentiment lexicon by incorporating information from various sources in this domain, such as sentiment labels and linguistic heuristics.

However, in many domains, the labeled data is usually in limited size and insufficient to extract accurate and robust sentiment information. In addition, since there are massive domains involved in online user generated content, it is expensive and time-consuming to manually annotate enough samples for each domain.

3.1 Optimization Method

In addition, the learning processes of sentiment classifiers in different domains are coupled together in our approach in order to exploit the sentiment relatedness among these domains. Thus, it is challenging to solve the optimization problem in our approach efficiently. In this work, this project introduce an accelerated algorithm based on FISTA to solve the model of our approach.

In addition, this project proposes a parallel algorithm based on ADMM to train sentiment classifiers for multiple domains in a parallel way, which can further improve the efficiency of our approach when domains to be analyzed are massive. Next this project will introduce them in detail.

3.2 Collaborative Online Multi Tasking

We study the problem of online multitask learning for solving multiple related classification tasks in parallel, aiming at classifying every sequence of data received by each task accurately and efficiently. One practical example of online multitask learning is the micro-blog sentiment detection on a group of users, which classifies micro-blog posts generated by each user into emotional or non-emotional categories.

First of all, to meet the critical requirements of online applications, a highly efficient and scalable classification solution that can make immediate predictions with low learning cost is needed. This requirement leaves conventional batch learning algorithms out of consideration.

Second, classical classification methods, be it batch or online, often encounter a dilemma when applied to a group of tasks, i.e., on one hand, a single classification model trained on the entire collection of data from all tasks may fail to capture characteristics of individual task; on the other hand, a model trained independently on individual tasks may suffer from insufficient training data. To overcome these challenges, in this paper, we propose a collaborative online multitask learning method, which learns a global model over the entire data of all tasks. At the same time, individual models for multiple related tasks are jointly inferred by leveraging the global model through a collaborative online-learning-approach.

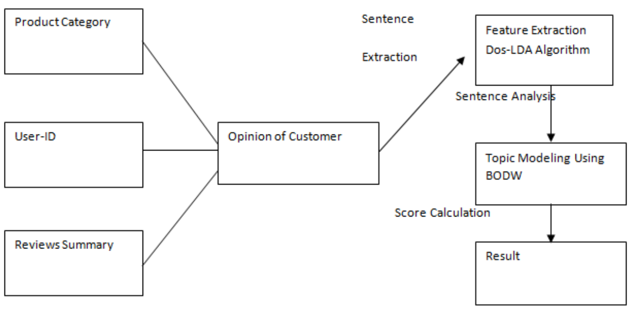

Fig 3.1 **system architecture**

3.3 Accelerated Algorithm

This project introduce the ENB(Enhanced Naïve Bayes) based accelerated algorithm for our approach which can be conducted on a single computing node. As mentioned before, the optimization problem in our approach is no smooth. Although this project can use sub gradient descent method to solve it, the convergence rate of sub gradient method is $O(\frac{1}{\sqrt{k}})$ and is far from satisfactory, where k is the number of iterations.

Different from gradient method and subgradient method where in each iteration the current solution is computed using the last solution, in FISTA the current solution is estimated using the last two solutions and the "momentum" between them is exploited to accelerate the optimization process. In each iteration of FISTA, two kinds of points are sequentially updated.

3.4 Active Learning

However, existing active learning work has mainly focused on training models for a single domain. In practical applications, it is common to simultaneously train classifiers for multiple domains. For example, some merchant web sites (like Amazon.com) may need a set of classifiers to predict the sentiment polarity of product reviews collected from various domains (e.g., electronics, books, shoes).

Though different domains have their own unique features, they may share some common latent features. If projects has apply active learning on each domain separately, some data instances selected from different domains knowledge due to the common features.

In our solution, a shared subspace is first learned to represent common latent features of different domains. By considering the common and the domain-specific features together, the model loss reduction induced by each data instance can be decomposed into a common part and a domain-specific part. In this way, the duplicate.

3.5 Customer Reviews

Merchants selling products on the Web often ask their customers to review the products that they have purchased and the associated services. As e-commerce is becoming more and more popular, the number of customer reviews that a product receives grows rapidly. For a popular product, the number of reviews can be in hundreds or even thousands. This makes it difficult for a potential customer to read them to make an informed decision on whether to purchase the product.

It also makes it difficult for the manufacturer of the product to keep track and to manage customer opinions. For the manufacturer, there are additional difficulties because many merchant sites may sell the same product and the manufacturer normally produces many kinds of products. In this research, we aim to mine and to summarize all the customer reviews of a product.

3 CONCLUSION

System presents a collaborative multi-domain sentiment classification approach. Approach can learn accurate sentiment classifiers for multiple domains simultaneously in a collaborative way and handle the problem of insufficient labeled data by exploiting the sentiment relatedness between different domains. The sentiment classifier of each domain is decomposed into two components, a global one and a domain-specific one.

The global model can capture the general sentiment knowledge shared by different domains and the domain-specific models are used to capture the specific sentiment expressions of each domain.

In addition, propose to incorporate the similarities between different domains into approach as regularization over the domain-specific sentiment classifiers to encourage the sharing of sentiment information between similar domains.

Moreover, to introduce an accelerated algorithm to solve the model of our approach efficiently, and propose a parallel algorithm to further improve its efficiency when domains to be analyzed are massive. Experimental results on benchmark datasets show that approach can effectively improve the performance of multi-domain sentiment classification, and significantly outperform baseline methods.

5 REFERENCES

1. Beck. A and Teboulle. M(2016), “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” SIAM J. Imaging Sci., vol. 2, no. 1, pp. 183–202..
2. Blitzer. J, Dredze. M, and Pereira. F(2014), “Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification,” in Proc. 45th Annu. Meeting Assoc. Comput. Linguistics, vol. 7, pp. 440–447.
3. Bollen. J, Mao. H, and Pepe. A(2013), “Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena,” in Proc. Int. AAAI Conf. Weblogs Social Media, pp. 17–21.
4. Boyd. S, Parikh. N, Chu. E, Peleato. B, and Eckstein. J(2015), “Distributed optimization and statistical learning via the alternating direction method of multipliers,” Found. Trends Mach. Learn., vol. 3, no. 1, pp. 1–122.
5. Cambria. E, Schuller. B, Xia. Y, and White. B(2016), “New avenues in knowledge bases for natural language processing,”Knowl.-Based Syst., vol. 108, no. C, pp.1–4.
6. Cambria. E(2016), “Affective computing and sentiment analysis,” IEEE Intell. Syst., vol. 31, no. 2, pp. 102–107.
7. Chen. T, Xu. R, He. Y, Xia. Y, and Wang. X(2016), “Learning user and product distributed representations using a sequence model for sentiment analysis,” IEEE Comput. Intell. Mag., vol. 11, no. 3, pp. 34–44.
8. Glorot. X, Bordes. A, and Bengio. Y(2013), “Domain adaptation for large-scale sentiment classification: A deep learning approach,” in Proc. 28th Int. Conf. Mach. Learn., pp. 513–520.
9. Go. A, Bhayani. R, and Huang. L(2015), “Twitter sentiment classification using distant supervision,” Stanford Univ., Stanford, CA, USA, Project Rep. CS224N, pp. 1–12.
10. Hu. M and Liu. B(2014), “Mining and summarizing customer reviews,” in Proc. 10th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, pp. 168–177.
11. Li. G, Hoi. S. C, Chang. K, Liu. W, and Jain. R(2014), “Collaborative online multitask learning,” IEEE Trans. Knowl. Data Eng., vol. 26, no. 8, pp. 1866–1876.
12. Li. L, Jin. X, Pan. S. J, and Sun. T(2012), “Multi-domain active learning for text classification,” in Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, pp. 1086–1094.
13. Li. S, Huang. C, and Zong. Q(2014), “Multi-domain sentiment classification with classifier combination,” J. Comput. Sci. Technol., vol. 26, no. 1, pp. 25–33.
14. Liu. B(2012), “Sentiment analysis and opinion mining,” Synthesis Lectures Human Language Technol., vol. 5, no. 1, pp. 1–167.
15. O’Connor. B, Balasubramanyan. R, Routledge. B. R(2013), and Smith. N. A, “From tweets to polls: Linking text sentiment to public opinion time series,” in Proc. Int. AAAI Conf. Weblogs Social Media, 2010, pp. 122–129.
16. Pang. B and Lee. L(2012), “Opinion mining and sentiment analysis,” Found. Trends Inf. Retrieval, vol. 2, no. 1/2, pp. 1–135.
17. Pang. B, Lee. L, and Vaithyanathan. S(2013), “ Sentiment classification using machine learning techniques,” in Proc. ACL Conf. Empirical Methods Natural Language Process., pp. 79–86.
18. Wu. F and Huang. Y(2015), “Collaborative multi-domain sentiment classification,” in Proc. IEEE Int. Conf. Data Mining, 2015, pp. 459–468.
19. Wu. F, Song. Y, and Huang. Y(2015), “Microblog sentiment classification with contextual knowledge regularization,” in Proc. 29th AAAI Conf. Artif. Intell., pp. 2332–2338.
20. Wu. Y,Liu. S, Yan. K, Liu. M, and Wu. F(2014), “OpinionFlow: Visual analysis of opinion diffusion on social media,” IEEE Trans. Vis. Comput. Graph., vol. 20, no. 12, pp. 1763–1772.