A Collaborative Model for Sentiment Analysis and Summarizing User Reviews Using Machine Learning and Data Mining Techniques

Jignesh Tanna, Deepak Mandal, Lalitkumar Lokhande, Shruti Bendale, Dr. Meenakshi Thalor

Information Technology
AISSMS Institute of Information Technology, Pune, India

Abstract:
The classification of emotions is a hot topic of research in the industrial and academic field. The main methods of classifying emotions are based on machine learning and treat the classification of emotions as a text classification problem. However, the classification of emotions is widely recognized as a highly domain-dependent task. The mood classifier trained in one domain may not work properly in another domain. A simple solution to this problem is to train a domain-specific emotion classifier for each domain. However, it is difficult to identify enough data for each domain because they are in a large amount. In addition, this method omits mood information in other domains. In this document, we propose to collaboratively train multi-domain sentiment classifiers based on learning multiple tasks. In particular, we divide the mood classifier into two domains in each domain, one generic and one domain specific. The general opinion classifier can capture the global opinion information and is trained in various areas to obtain a better generalizability. The domain-specific opinion classifier is trained using data labeled in a domain to capture the domain-specific opinion information. In addition, we examine two types of relationships between domains, one based on the text content and the other based on the distribution of mood words. We create a domain similarity diagram using domain relationships and encode it in our approach as regularization for domain-specific opinion classifiers. In addition, we incorporate the knowledge of the feeling extracted from the lexicons of feeling to more accurately train the general classifier of emotions. In addition, we present an accelerated optimization algorithm for efficiently training sentiment classifiers. Experimental results in two sets of reference opinion data show that our method can significantly and consistently exceed basic methods.

Keywords: Data mining, Web mining, Information Search and Retrieval, Sentimental classification, multi-task learning, sentimental analysis, Natural Language Processing, Deep Learning, Question Answering System, Automatic Evaluation, Sequence Matching

I. INTRODUCTION
Throughout Web 2.0, the second phase of Internet development, static web pages become dynamic or user-generated content and social network growth. The benefits of Web 2.0 are that they are available anytime, anywhere, a variety of media, ease of use, active student participation in building knowledge, and creating dynamic learning communities The editor, every issue that can be tracked, is simple to use and offers a discussion in real time.
Social Networking-
Is the use of Internet-based social media programs to connect with friends, family, classmates, customers and customers. It can be for social, commercial, or both via websites such as Facebook, Twitter, Linked In, classmates.com, and Yelp. It is an important destination for sellers who want to attract users. The benefits of social networking are global connectivity, shared interests, real-time information sharing, and targeted advertising. The most important social networks are Twitter, Facebook, Linked In, Google+, You Tube, Instagram and Snap Chat, and so on

Twitter-
Send a message, picture, etc. to the social networking service twitter. A social networking site where the user can post short messages that are visible to other users. These messages are called tweets and can be up to 140 characters long. It was founded in 2006. As of 2008, Twitter had an estimated 4 to 5 million users and was the third most popular social networking site after Facebook and MySpace. Tweets are news, pictures, etc. posted on Twitter.

Sentiment Analysis-
It is also referred to as Opinion Mining and refers to the use of Natural Language Processing (NLP), text analysis, computational linguistics and biometrics for the systematic identification, extraction, quantification and investigation of the objective states of subjective information. It is widely applied to the voice of customer materials such as reviews and survey responses, social and online media, and healthcare materials for applications ranging from marketing to customer service to clinical medicine. Mentioned is the process of determining the emotional tone behind the word string used to convey an understanding of the attitudes, opinions and emotions expressed in an online system. The proposal is about learning to insert certain sentiment words. Only the context of the words is used and the feeling of the texts is ignored. Words are mapped with a similar context but with an opposite feeling. To get more information about incorporating feelings into the number of neural networks and loss functions, they can also be used as features of natural words. The analysis of feelings can be applied to the analysis of feelings at the word level, the classification of feelings at the level of prayer, and the construction of emotional encyclopedias. The word representation represents aspects of the meaning of the words. Each word is a continuous vector of small dimension and real value, referred to as natural language processing tasks in word embedding, machine translation, syntactic analysis, answering questions, discourse analysis.

Embedding of learning feelings that encode the feeling of texts in the continuous representation of words. Develop a series of neural networks with the loss of adjustment functions to learn how to incorporate feelings. It is used to tweet with positive and negative emoticons, e.g. B. with remotely controlled body without manual annotation. Review the effectiveness of incorporating feelings when applied to three tasks of emotion analysis.

The empirical test results show that the inclusion of feelings outweighs the contextual inclusion in several reference records of the opinion classifier of these tasks. Therefore, the ant-grouping algorithm is used to identify aspects. Take similar sentences and group them. Then extract another aspect of the objective opinion object. In the sentiment analysis, the different sentences in a product evaluation refer to the different aspects of the products being evaluated. Growth of hierarchical maps of self-organization for the classification of revision sentences. In this way, we can see if the various aspects, such as the example product mentioned with positive feelings, are negative in the scores .

LITERATURE SURVEY
In this paper, we present the top-20 cited papers from Google Scholar and Scopus and a taxonomy of research topics. In recent years, sentiment analysis has shifted from analyzing online product reviews to social media texts from Twitter and Facebook. Many topics beyond product reviews like stock markets, elections, disasters, medicine, software engineering, and cyberbullying extend the utilization of sentiment analysis.

2] Poorv Agarwal Boyi Xie Ilia Vovsha Owen Rambow Rebecca Passonneau, “Sentiment Analysis of Twitter Data”
In this paper, we examine sentiment analysis on Twitter data. The contributions of this paper are: (1) they introduce POS-specific prior polarity features. (2) they explore the use of a tree kernel to obviate the need for tedious feature engineering. The new features (in conjunction with previously proposed features) and the tree kernel perform approximately at the same level, both outperforming the state-of-the-art baseline.

In this system, the central aim of this research paper is to perform sentiment analysis on movie review data. They have proposed the Senti-lexical algorithm to find the polarity of a review as positive, negative, or neutral. We have also proposed a method to handle words which have a negation effect on the reviews and the role of emoticons is also discussed.

4] Xiaomei Zou, Jing Yang, Jianpei Zhang, “Micro blog sentiment analysis using social and topic context”
In this paper, they propose a new method combining social context and topic context to analyze microblog sentiment. In particular, different from previous work using direct user relations, we introduce structure similarity context into social contexts and propose a method to measure structure similarity. In addition, we also introduce topic context to model the semantic relations between microblogs. Social context and topic context are combined by the Laplacian matrix of the graph built by these contexts and Laplacian regularization are added into the microblog sentiment analysis model. Experimental results on two real Twitter datasets demonstrate that our proposed model can outperform baseline methods consistently and significantly.

This paper presents statistics on the evolution of sentiment analysis. What kind of topics have been discussed? How has their popularity changed over time? Who have been the leading researchers? Answers to these questions are provided by statistical analysis on keywords and by applying Latent Dirichlet Allocation to the titles and abstracts of the publications. The aim of this paper is to provide background information on the big picture of semantic analysis and its development over time.

In the specific field of sentiment analysis tasks, traditional machine learning models lack effective attention to certain features. To address this issue, we extracted and classified user emotion features based on the LSTM network model and applied them to perinatal depression screening scenarios. We use emoticons as feature extraction and modeling for document-level sentiment analysis in specific regions and achieved good results. The results were basically consistent with the findings of the Edinburgh Postnatal Depression Scale. This method greatly shortens the screening time and reduces
the doctor-patient communication costs. It has positive significance for specific areas of sentiment classification tasks and provides references for document-level sentiment analysis.

7] Imane Alaoui, Youssef Gahi, Rochdi Messoussi, Youness Chaabi, Alexis Todosk off and Abdessamad Kobi,
“A novel adaptable approach for sentiment analysis on big social data”
In that the proposed approach consists of first constructing a dynamic dictionary of words’ polarity based on a selected set of hashtags related to a given topic, then, classifying the tweets under several classes by introducing new features that strongly fine-tune the polarity degree of a post. To validate our approach, we classified the tweets related to the 2016 US election. The results of prototype tests have performed a good accuracy in detecting positive and negative classes and their sub-classes.

8] Suhaila Zainudin, Zohreh Madhoushi, Abdul Razak Hamdan,
“Sentiment analysis techniques in recent works”
This survey aims to categorize SA techniques in general, without focusing on specific level or task. And also to review the main research problems in recent articles presented in this field. We found that machine learning-based techniques including supervised learning, unsupervised learning and semi-supervised learning techniques, Lexicon-based techniques and hybrid techniques are the most frequent techniques used. The open problems are that recent techniques are still unable to work well in different domain; sentiment classification based on insufficient labeled data is still a challenging problem; there is lack of SA research in languages other than English; and existing techniques are still unable to deal with complex sentences that requires more than sentiment words and simple parsing.

9] Shuhaida Mohamed Shuhidan, Saidatul Rahah Hamidi, Soheil Kazemian, Shamila Mohamed Shuhidan and Maizatul Akmar Ismai,
“Sentiment Analysis for Financial News Headlines using Machine Learning Algorithm”
This study consists of several phases in pre-processing such as extract data, stop word removal, and stemming to clean the dataset and make it as data preparation before performing the sentiment analysis with the selected machine learning algorithms. In the stop word removal, tm package in R is used to clean the dataset while for stemming process, Snowball stemmer is used to set the data to its root word. Sample outcomes of analysis are explained for both algorithms. The conclusion describes the summation of the study and future works.

10] Ying Fang, Hai Tan, and Jun Zhan,
“Multi-Strategy Sentiment Analysis of Consumer Reviews Based on Semantic Fuzziness”
In this paper Sentiment analysis mines opinions at word, sentence, and document levels, and gives sentiment polarities and strengths of articles. As known, the opinions of consumers are expressed in sentiment Chinese phrases. But due to the fuzziness of Chinese characters, traditional machine learning techniques can not represent the opinion of articles very well. In this paper, we propose a multi-strategy sentiment analysis method with semantic fuzziness to solve the problem. The results show that this hybrid sentiment analysis method can achieve a good level of effectiveness.

“Sentiment Analysis of Big Data: Methods, Applications, and Open Challenges”
The application of opinion mining and sentiment analysis (OMSA) in the era of big data have been used a useful way in categorizing the opinion into different sentiment and in general evaluating the mood of the public. Moreover, different techniques of OMSA have been developed over the years in different data sets and applied to various experimental settings. In this regard, this paper presents a comprehensive systematic literature review, aims to discuss both technical aspect of OMSA (techniques and types) and non-technical aspect in the form of application areas are discussed.
Furthermore, this paper also highlighted both technical aspects of OMSA in the form of challenges in the development of its technique and non-technical challenges mainly based on its application. These challenges are presented as a future direction for research.

12] Ali Hasan, Sana Moin, Ahmad Karim and Shahaboddin Shamshirb, “Machine Learning-Based Sentiment Analysis for Twitter Accounts” In this paper we study the use of various machine-learning techniques and tools for sentiment analysis during elections, there is a dire need for a state-of-the-art approach. To deal with these challenges, the contribution of this paper includes the adoption of a hybrid approach that involves a sentiment analyzer that includes machine learning. Moreover, this paper also provides a comparison of techniques of sentiment analysis in the analysis of political views by applying supervised machine-learning algorithms such as Naïve Bayes and support vector machines (SVM).

II. PROBLEM STATEMENT
Users’ informal conversations like chats, post, tags, comments etc. on online social media e.g. Twitter, Facebook shed light into their educational experiences such as their opinions, feelings, and all related concerns about the learning process activities. Data from such un-instrumented environments can provide valuable knowledge to inform user communication process. Analyzing such data, it can be challenging task for researcher.

III. EXISTING SYSTEM
The existing system should introduce the background of the word embedding. Then, the methodology is introduced to learn how to incorporate feelings into the next segment. The use of incorporating feelings in three applications occurs in the next half, i.e, the analysis of feelings at the word level, the classification of feelings at the prayer level, and the lexicon of feeling construction. There are several models: predictive model, namely classification model, hybrid prediction model, hybrid classification model.

This section also introduces the emotion-learning methods that describe context-based neural network methods for learning how to incorporate words, and introduces our extension to capture the polarity of sentences before presenting hybrid models that encode both emotionlevel information as well as the context and then describe the integration of information at the word level to embed learning. Most existing studies that encode only word contexts in word inlays. Embed certain words of feelings that facilitate the ability to embed words to capture similarities of words in terms of the semantics of feelings. Several neural networks encode the context and the information on a sentimental level. Terms such as "good" and "bad" can be separated in the information on the level of opinion.

Disadvantages
• Dependency on transactional data
• No any identification strategy

V. PROPOSED METHOD
The regularized multi-task learning method proposed in. The model of each task is constrained to be similar with their average model. Our proposed collaborative multi-domain sentiment classification approaches with squared loss, hinge loss, and log loss respectively.
A novel domain similarity measure based on sentiment word distributions is proposed. We formulate the model of our approach into a convex optimization problem. Moreover, we 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.
Collaborative multi-domain sentiment classification approach. Our 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. We propose to extract domain-specific sentiment knowledge from both labeled and unlabeled samples, and use it to enhance the learning of the domain specific sentiment classifiers. We propose to use the prior general sentiment knowledge in general-purpose sentiment lexicons to guide the learning of the global sentiment classifier. In addition, we propose to incorporate the similarities between different domains into our approach as regularization over the domain-specific sentiment classifiers to encourage the sharing of sentiment information between similar domains.

Advantages of the Proposed System:
- Speeding up the learning process by training sentiment classifiers for multiple domains in parallel at different computing nodes.
- A parallel algorithm to further improve its efficiency when domains to be analyzed are massive.

System Architecture Diagram

VI. CONCLUSION
Get to know the specific word inlays of most existing studies that encode only word contexts in word embedding, the feeling of texts to facilitate the ability of word embedding to capture word similarity in terms of the semantics of feelings. As a result, words having similar contexts but having opposite feeling polarity terms such as "good" and "bad" can be separated into the space of inclusion of feelings and multiple neural networks to effectively simultaneously encode context information and feeling level in the inclusion of words in some way. The effectiveness of incorporating feelings is empirically verified in three emotional analysis tasks. In analyzing word-level feelings, we show that incorporating emotions is useful in detecting similarities between feelings. In classifying feelings at the prayer level, the inclusion of feelings is useful for capturing discriminating features in order to predict the feeling of the sentences. Lexical-level tasks, such as the construction of the feeling dictionary, show that inclusion of emotions is useful in measuring the similarities between words. Hybrid models that capture both contextual and mood information are the best in all three tasks.

VII. FUTURE WORK
Further, we can widespread the RNN & RNN-Naïve Hybrid algorithm to a broad range of sentiment analysis tasks. We also plan to consider more complex polarity shift patterns such as transitional subjunctive and sentiment-inconsistent sentences in creating reversed reviews.

ACKNOWLEDGMENT
Thanks to our Guide Dr. M.A. Thalor, AISSMS Institute of Information Technology, Pune for providing resources and helping us in all possible ways. We also thank readers of this journal for.
showing interest in this topic and contributing towards enhancement of this topic as well.

REFERENCES


