A ES

ISSN: 2096-3246
Volume 54, Issue 02, May, 2022

https://doi.org/10.5281/zenodo.12707402

COLLABORATIVE MULTI-DOMAIN SENTIMENT CLASSIFICATION APPROACH FOR SOCIAL MEDIA

SREEKANTAM VASUDHA,DIGALA RAGHAVA RAJU,K BALAJI SUNIL CHANDRA

Assistant Professor[1,2,3],

vasudhasvit87@gmail.com, raghava.digala@gmail.com,
hod.cse@svitatp.ac.in,

department of CSE, Sri Venkateswara Institute of Technology,
N. H 44, Hampapuram, Rapthadu, Anantapuramu, Andhra Pradesh 515722

Keywords:

There is an issue with data route revocation, TDMA, or medium access control.

ABSTRACT

Our strategy is based on the central principle of sentiment analysis, which is the study of human health as it relates to social media. It is well-known that current sentiment categorization is an issue that is very domain-dependent. So, it accurately predicts a person's depression rate but offers an extremely low estimate. Using a collaborative multi domain sentiment classifier, which has the benefit of more precisely determining a person's depressive condition, may fix the issue. Our goal is to use multi-task learning as a collaborative framework to train sentiment classifiers for various domains. Its many applications include marketing research, political campaigns, brand messaging, and consumer feedback gathering. Throughout the subject review The Bag of Words (BODW) approach is crucial. The most crucial words in a paper could stand for an opinion or a fact about the subject. Both the fact and the feeling stand for the objective and subjective labels, respectively. From a manuscript, the system extracts a bag of discriminative words using factors that may be subjectively or objectively selected. To remove discriminatory terms from a text, LDA and regression methods are used. Recommendations are made by comparing subjective scores with respect to the user's query, which is based on both objective and subjective analysis. One method that does this is support vector machine. Naïve Bayes, a classification algorithm, can simultaneously detect changes in sentiment words when a new dataset is applied and classify them as positive or negative, all without linguistic resources. We may use this data to determine who is suffering from depression by calculating their score. With the support of official authorization to access social media data, this may be accomplished.

ISSN: 2096-3246
Volume 54, Issue 02, May, 2022
https://doi.org/10.5281/zenodo.12707402

Introduction

People from all across the globe express their thoughts, ideas, and sentiments on social media platforms such as Twitter and Facebook. Emotional text analysis and classification, however, is a formidable task that represents a sophisticated version of Sentiment Analysis. A system for dividing text into "Happy" and "Sad" sentiment categories is shown here. Our programme successfully extracts these emotions from text by using two distinct ways. Emotional indicators, degree words and negations, Parts of Speech, and other grammatical analyses are used in the first technique, which is based on Natural Language Processing. The second method makes use of techniques for machine learning classification. In addition, we have developed a system to automatically generate training sets, doing away with the requirement for human annotation of massive datasets. Our team has also amassed a sizable collection of emoticons and the levels of intensity associated with them. Results from tests demonstrate that our model outperforms the state-of-the-art in Twitter tweet classification. Crisis management, social publicity, removing client interest, personalised recommendations, and purchaser relation management may all benefit from analysing consumer produced satisfaction feelings. There are several uses for mining the sentiment information in huge amounts of user-generated material, which may serve to gauge public opinion on a variety of subjects (e.g., goods, brands, catastrophes, events, celebrities, and so on). It is also beneficial to classify the feelings of large microblog postings as an alternative to or in addition to the time-consuming and costly conventional polls. Therefore, emotion categorization is a highly sought-after area of study in academia and business. Sentiment organisation is seen as a passage classification challenge in several majority sentiment investigation methodologies. To predict the sentiments of unseen texts, sentiment classifiers are trained on labelled datasets using supervised machine learning methods like SVM and Logistic Regression. Various techniques have been used to examine the

sentiments of product reviews, micro blogs and so on. On the other hand, sentiment classification is widely recognized as a domain- dependent problem. This be dissimilar domains present are different response words, and the equal

.

II. PROPOSED SYSTEM

Two kinds of data are combined to extract domain-specific sentiment knowledge for each domain. The first kind of data is the labeled samples, which are associated with sentiment labels and can be used to infer domain- specific sentiment expressions directly. A common observation in sentiment analysis field is that the words occurring more frequently in happy samples than sad samples usually tend to convey positive sentiment orientations, and vice versa. Thus, we can propagate the sentiment labels from documents/sentences to words to extract the domain-specific sentiment expressions. Several preprocessing steps were taken before experiments. Words were converted to lower cases and stop words were removed. In this paper, we propose to extract the initial sentiment scores of words based on their distribution differences in happy and sad samples.

Advantages:

- Can outperform multiple task learning
- Less time consuming
- Uses less hardware and software
- Find the depression state of a person by using multi-domain

III. PROJECT DESCRIPTION

Implementation is the stage of the project when the theoretical design is turned out into

ISSN: 2096-3246
Volume 54, Issue 02, May, 2022
https://doi.org/10.5281/zenodo.12707402

a working system. Thus it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective.

Domain Dataset:

In domain Datasets we have included two datasets one is from facebook and another is from whatsapp. This is loaded with the neggets process as a document. The datasets will be loaded as a conversation part as this is a chats, this will be easy for us to find who have conversed by having their name. This datasets will be loaded for further process of preprocessing where we will remove all the unwanted words.

PREPROCESSING:

LDA and BoDW

Articles and punctuation marks are examples of stopping words that will be removed during the preprocessing step. With the assistance of LDA and BoDW, this procedure will be completed. Since we had provided halting words in one document and slang in another—slang being a contraction of words that would not aid in determining a person's depressed state—we eliminated the slang. With these two texts in hand and the algorithms LDA and BoDW running the show, we can filter out any irrelevant phrases that may obscure a person's sadness. Once these methods are used, we will have a compressed text that is both too short and too exact for us to easily locate the words. Two techniques, the Global model and the Domain Similarity, will be available after this preprocessing step. them of these options are available to us, and we have made use of them in our project. LDA: Topic modelling is one of the oldest and most well-known challenges in NLP and ML. Blei, Ng, and Jordan's Latent Dirichlet Allocation is one of the most effective generative latent topic models, and we introduce it in this chapter. When it comes to statistical analysis, Latent Dirichlet Allocation (LDA) is a great technique for document collections and other types of discrete data. To be more precise, LDA is a graphical model that does not rely on labels to find hidden themes in data. For the purpose of filtering information from big datasets, the "many users" required for collaborative processes are difficult to exist from the start, and this is precisely why LDA makes this model superior to collaborative models. BoDW: Bag of Words is a crucial tool for subject analysis. The most crucial words in a paper could stand for an opinion or a fact about the subject. The fact stands for the neutral terms, and the feeling for personal designations. Depending on the subject, the definition of a term could change. The degree to which each word in a text represents a subject varies. A word's discriminatory power, both subjectively and objectively, changes depending on the context. From a manuscript, the system extracts a bag of discriminative words using factors that may be subjectively or objectively selected. To remove discriminatory terms from a text, LDA and regression methods are used. Analysing the user's subjective score in relation to their query allows for the generation of suggestions based on both objective and subjective analysis.

Global model

In global model, the preprocessed two domain will be globally made into consideration for the sentiment analysis. As the name suggest global refers to is one which predicts fracture by considering the entire body: example is a classic LEFM model of a body containing a crack - failure occurs at a critical value of the stress intensity, k, which is a function of the geometry and loading of the whole body. Here, we have applied Naive's Bayes algorithm to have a better results of the sentence which is present is happy or sad or normal. This algorithm will help us to

ISSN: 2096-3246
Volume 54, Issue 02, May, 2022
https://doi.org/10.5281/zenodo.12707402

do this process in a simple way. From global model we could able to achieve Domain Similarity. For this we can apply SVM(Support Vector Machine) algorithm.

Navie Bayes Classifier:

Naive Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems. It is mainly used in text classification that includes a high-dimensional training dataset. Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions. It is a probabilistic classifier, which means it predicts on the basis of the probability of an object. Some popular examples of Naïve Bayes Algorithm are spam filtration, Sentimental analysis, and classifying articles. Naive: It is called Naïve because it assumes that the occurrence of a certain feature is independent of the occurrence of other features. Such as if the fruit is identified on the bases of color, shape, and taste, then red, spherical, and sweet fruit is recognized as an apple. Hence each feature individually contributes to identify that it is an apple without depending on each other. Bayes: It is called Bayes

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

because it depends on the principle of Bayes' Theorem.
P(A|B) is Posterior probability: Probability of hypothesis A on the observed event B. P(B|A) is Likelihood probability: Probability of the evidence given that the probability of a hypothesis is true.
P(A) is Prior Probability: Probability of hypothesis before observing
the evidence P(B) is Marginal Probability: Probability of Evidence.

Domain Similarity:

One of the most challenging problems in natural language processing (NLP) is to figure out whether two texts belong to the same domain or not. This problem has applications in vast areas such as smart advertising robots, which look for semantically related pages to show an ad, or robots, which are supposed to collect news pertinent to a specific topic. Generally speaking, a measure for deciding how close domains of two texts are can be used as a metric for text distance in text classification tasks. In this research, we strived to develop a system for a task which is a combination of two famous tasks in NLP, namely semantic textual similarity (STS) and text classification. In this task, two texts are compared to each other like the STS task, but rather than their semantics, their classes are the basis of this comparison.

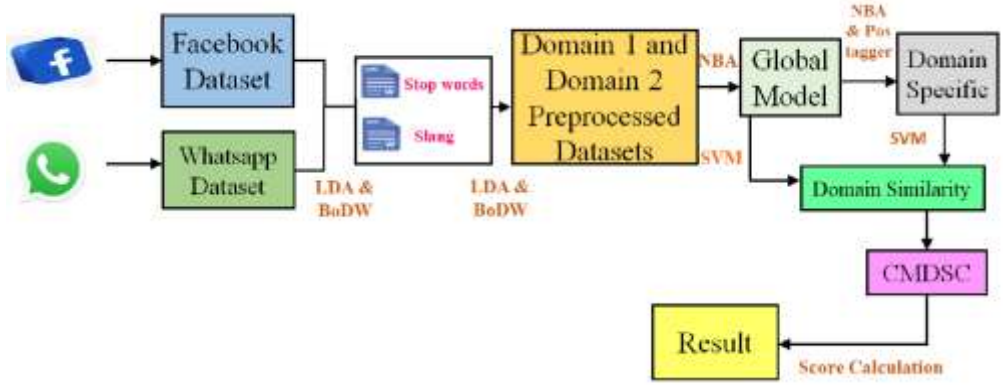
Collaborative Multi-Domain Sentiment Classification

The mainstream sentiment classification methods are based on machine learning and treat sentiment classification as a text classification problem. However, sentiment classification is widely recognized as a highly domain-dependent task. The sentiment classifier trained in one domain may not perform well in another domain. A simple solution to this problem is training a domain-specific sentiment classifier for each domain. However, it is difficult to label enough data for every domain since they are in a large quantity. In addition, this method omits the sentiment information in other domains. In this paper, we propose to train sentiment classifiers for multiple domains in a collaborative way based on multi-task learning. The general sentiment classifier can capture the global sentiment information and is trained across various domains to

ISSN: 2096-3246
Volume 54, Issue 02, May, 2022
https://doi.org/10.5281/zenodo.12707402

obtain better generalization ability. The domain- specific sentiment classifier is trained using the labeled data in one domain to capture the domain-specific sentiment information. In addition, we explore two kinds of relations between domains, one based on textual content and the one based on sentiment word distribution. We build a domain similarity graph using domain relations and encode it into our approach as regularization over the domain-specific sentiment classifiers. Besides, we incorporate the sentiment knowledge extracted from sentiment lexicons to help train the general sentiment classifier more accurately. Moreover, we introduce an accelerated optimization algorithm to train the sentiment classifiers efficiently. Experimental results on two benchmark sentiment datasets show that our method can outperform baseline methods significantly and consistently.

IV. SYSTEM MODEL



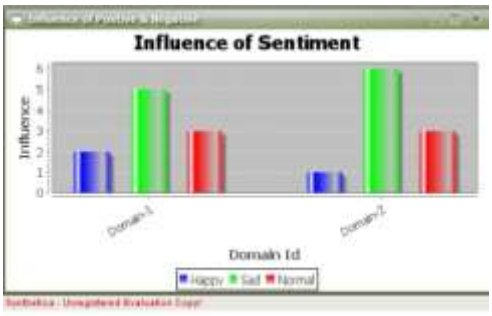
RESULT AND DISCUSSION DATASETS AND EXPERIMENTAL SETTINGS

Two benchmark multi-domain sentiment datasets were used in our experiments. The first one is the famous sentiment dataset1 (denoted as facebook), which was collected by the users. and includes two domains, i.e., Facebook and Whatsapp. It is widely used in multi-domain and cross-domain sentiment analysis fields. In each domain, 1,000 positive and 1,000 negative reviews are included.

COMPARISON OF DOMAIN SIMILARITY MEASURES

In this section we conducted experiments to figure out which one of the two domain similarity measures introduced is more suitable for multi-domain sentiment classification task. The experimental results on facebook dataset are the results on whatsapp dataset show similar parents. Hinge loss was used in our approach in these experiments.

ISSN: 2096-3246
Volume 54, Issue 02, May, 2022
https://doi.org/10.5281/zenodo.12707402



The performance of our approach with different kinds of domain similarity. NoSim(red), ContentSim(blue), and SentiSim(green) represent the performance of our approach with no domain similarity, with textual content based domain similarity, and with sentiment expression based domain similarity respectively. The difference between SentiSim-Initial and SentiSim-Prop is that the former is based on the initial sentiment scores extracted from labeled samples, and the latter is based on the sentiment scores after propagation. we can see that the performance of our collaborative multi-domain sentiment classification approach with sentiment expression based domain similarity is better than that with textual content based domain similarity. This result indicates that the domain similarity based on sentiment expressions can better measure the sentiment relatedness between different domains than that based on textual content in multi-domain sentiment classification task.

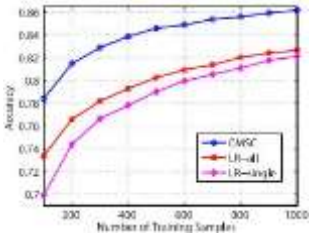
INFLUENCE OF TRAINING DATA SIZE

In this section, we conducted experiments to explore the influence of training data size on the performance of our approach. We want to verify whether our approach can alleviate the requirement for labeled samples by training sentiment classifiers for multiple domains collaboratively. In our experiments, we varied the number of training samples in each domain from 100 to 1,000, with a step size of 100. The loss function used in our approach is log loss. Thus our approach significantly outperforms LR-all. Although LR-single also builds domain-specific sentiment classifiers for each domain by training on the labeled samples of this domain, our approach outperforms it signifi- cantly. In addition, the performance improvement of our approach over LR-single method is more significant when the labeled data in each domain is scarce.

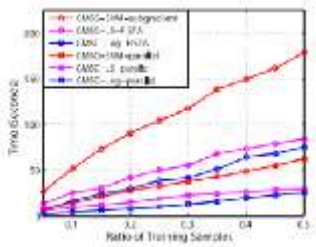
TIME EFFICIENCY

We conducted several experiments to explore the time complexity of our approach. All experiments were conducted on a desktop computer with Intel Core i7 CPU (3.4 GHz) and 16 GB RAM. The single-node version FISTA-based accelerated algorithm was conducted on a single core of this machine, and the ADMM-based parallel algorithm was distributed across the 4 cores of this machine. The experiments were conducted on the Amazon-21 dataset. In each experiment we randomly selected r of the labeled samples in each domain for training.

ISSN: 2096-3246
Volume 54, Issue 02, May, 2022
https://doi.org/10.5281/zenodo.12707402



The average performance of our approach and baseline methods on the two domains of dataset with different numbers of training samples. CMSC represents our collaborative multi-domain sentiment classification approach. LR-all and LR-single are Logistic Regression sentiment classifiers trained on all labeled samples and single-domain labeled samples, respectively. We can see that the running time of our approach with different kinds of loss functions is approximately linear with respect to size of the training data.



This result validates our analysis of the time complexity. Besides, our approach with log loss (CMSC-Log) and squared loss (CMSC-LS) runs much faster than that with hinge loss (CMSC-SVM). It validates the usefulness of the accelerated algorithm based on FISTA in improving the efficiency of our approach. In addition, the running time of the parallel algorithm is significantly less It validates the effectiveness of our parallel algorithm in speeding up the learning process by training sentiment classifiers for multiple domains in parallel at different than that of single-node version optimization algorithm.

v. 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. 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. Besides, propose to use the prior general sentiment knowledge in general-purpose sentiment lexicons to guide the learning of the global sentiment classifier. In addition, propose to incorporate the similarities between different domains into approach as regularization over the domain-specific sentiment classifiers

ISSN: 2096-3246
Volume 54, Issue 02, May, 2022

https://doi.org/10.5281/zenodo.12707402

to encourage the sharing of sentiment information between similar domains. Formulate the model of approach into a convex optimization problem. 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

VI. REFERENCES

[1]."A systematic literature review of sentiment analysis," International Journal of Computer Science and Engineering, volume 5, issue 4, pages 22–28, 2017, by J. Kaur, S. S. Sehra, and S. K. Sehra. [2]"A survey of multimodal sentiment analysis," Image Vis. Comput., vol. 65, pp. 3–14, Sep. 2017, by M. Soleymani, D. Garcia, B. Jou, B. Schuller, S.–F. Chang, and M. Pantic. Alonso, C. Gómez-Rodríguez, and D. Vilares published an article titled "Supervised sentence analysis in multilingual environments." Computer Science and Operations Management, volume 53, issue 3, pages 595-607, May 2017. "Enhancing deep learning sentiment analysis with ensemble techniques in social applications," by O. Araque, I. Corcuera-Platas, J. F. Sánchez-Rada, and C. A. Iglesias Publication date: July 2017, volume 77, pages 236–246, DOI: 10.1016/j.eswa.2017.02.002. In the proceedings of the 50th annual meeting of the Association for Computing Linguistics, volume 2, held on Jeju Island, South Korea in 2012, S. Wang and C. D. Manning discuss "Baselines and bigrams: Simple, good sentiment and topic classification" (pp. 90-94). [On the web]. This resource may be accessed at: http://dl.acm.org/citation.cfm?id=2390665.2390688. In the 2014 Proc. Conf. Empirical Methods Natural Language Processing (EMNLP), L. Zhao, M. Huang, H. Chen, J. Cheng, and X. Zhu presented their work on "Clustering aspect-related phrases by leveraging sentiment distribution consistency," which was published on pages 1614–1623. [On the web]. Visit: http://emnlp2014.org/papers/ pdf/EMNLP2014169.pdf to see the whole version. Proc. 15th International Symposium on Parallel Computing (ISPDC), 2016, pp. 230-233, doi:10.1109/ispdc.2016.39, "Comparison of text sentiment analysis based on machine learning" [7] by X. Zhang and X. Zheng. The authors of the article "Sentiment classification of online reviews to travel destinations by supervised machine learning approaches" (Q. Ye, Z. Zhang, and R. Law, 2009) published in Expert Systems, Artificial Intelligence and Applications, did not include a citation for their work. [9]In the proceedings of the 5th International Conference on Fuzzy Neuro-Computing (FANCCO), 2015, S. Mahalakshmi and E. Sivasankar discuss "Cross domain sentiment analysis using different machine learning techniques," which can be found on pages 77–87. The paper "Evaluating cross domain sentiment analysis using supervised machine learning techniques" was presented at the 2017 IntelliSys Conference and can be found in the proceedings (pp. 689-696). The authors are A. A. Aziz, A. Starkey, and M. C. Bannerman. The DOI for the paper is 10.1109/ intellisys.2017.8324369. Morgan & Claypool Publishers, B. Liu [11]. Published in 2012 by Morgan & Claypool. Topologically accurate feature maps self-organize, [12]T. Kohonen Publication date: 1982, volume 43, issue 1, pages 59–69, DOI: 10.1007/ bf00337288 in Biol. Cybern. *"Techniques and applications for sentiment analysis," by R. Feldman Volume 56, Issue 4, Page 82, April 2013, Journal of the Association for Computing Machinery, DOI: 10.1145/2436256.2436274. "Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN," published in Expert Systems Technology and Applications in April 2017, doi:10.1016/j.eswa.2016.10.065, was written by T. Chen, R. Xu, Y. He, and X. Wang. *"An unsupervised aspect detection model for sentiment analysis of reviews" (Proc. Natural Lang. Process. Inf. Syst., 2013, pp. 140-151), written by A. Bagheri, M. Saraee, and F. de Jong. [16]"Implicit aspect extraction in sentiment analysis: Review, taxonomy, oppportunities, and open

ISSN: 2096-3246
Volume 54, Issue 02, May, 2022
https://doi.org/10.5281/zenodo.12707402

challenges," Inf. Process. Manage., vol. 54, no. 4, pp. 545-563, Jul. 2018, doi:10.1016/j.ipm.2018.03.008. The authors are M. Tubishat, N. Idris, and M. A. Abushariah in 2018. The following is an excerpt from an article published in the Journal of Computing Linguistics in June 2011 with the DOI:10.1162/coli_a_00049: "Lexicon-based methods for sentiment analysis" by M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede. "Predicting the semantic ori-entation of adjectives," in Proc. 8th Conf. Eurochapter Assoc. Computer Linguistics, 1997, pp. 174–181, doi:10.3115/979617.979640, written by V. Hatzivassiloglou and K. R. McKeown.