AN EFFICIENT SENTIMENT ANALYSIS ON LIVE TWITTER DATA BASED ON CLASSIFICATION SYSTEM

Mrs. R.K. Santhia M.E,
Assistant Professor,
Department of Computer Science and Engineering,
Manakula Vinayagar Institute of Technology,
Puducherry.

Alamelumangai.C.T
Department of Computer Science and Engineering,
Manakula Vinayagar Institute of Technology,
Puducherry.

Lavanya.P
Department of Computer Science and Engineering,
Manakula vinayagar Institute of Technology,
Puducherry.

Priyadharshini.R
Department of Computer Science and Engineering,
Manakula Vinayagar Institute of Technology,
Puducherry.

Abstract: Each and every person life is filled with perspective and opinions. People like to share their emotions and opinions at every situation but social media is one of the major and easy way to share our feelings. Nowadays people not only comment on the extant information, pages of bookmarks and ratings are provided but they also share their ideas, news and knowledge with the community at huge. In Sentiment analysis we use natural language processing and information to extract writer's comments or reviews. In this paper we use Data text mining and hybrid approach of KNN Algorithm and Naive Bayes Algorithm to obtain the sentiments of people's thought in the world on Twitter. Social media is a computer technology that facilitates the sharing of ideas, thoughts and information. In the field of sentiment analysis there has been lot of work in twitter data. This survey focuses mainly on twitter data in sentiment analysis which is useful to examine the tweets of the data where thoughts are highly unorganized, composite and are either positive or negative, or neutral in some manner. We exhibit that recurrent neural networks, specifically character-occupying ones, can upgrade over latent semantics indexing models and bag-of words. Considering these models of the transfer capabilities are worse, the newly proposed training heuristics provides a unison model with performance equivalent to that of the three single models.

Keywords:
Sentiment analysis, Machine Learning, Natural Language Processing, Opinion Finder, convolution Neural Networks, Recursive Neural Network.

I. INTRODUCTION

Human life is filled with emotions and opinions. People like to share their emotions and opinions at every situation but social media is one of the major path and it is easy to share our feelings. In this generation many of them not only comment on the actual information, pages of bookmark and provided ratings but they also share their thoughts, news and knowledge with the community at large. In this step, the community not only being a reader but also becomes a writer. The actual mediums like Blogs, Wikis, Forums and Social Networks in which the individual can post message, provide opinions and obtain feedback from an another individual on various topics, pasturing from politics and health to product reviews and travelling. Recently, this area can be focused by many researchers. They are trying to fetch opinion to analyze and summarize the thoughts convey automatically with computers. While composing it, the writer wanted to express the emotion which can be emotion recognition algorithm. We need to define a set of basic emotions to handle this problem as a special case of text classifications. There is no single, standard set of basics emotions which have been studied by psychologists. Hence we assured to work with these three classifications that are most favorable, and also been used before by the researchers from computational linguistics and natural language processing (NLP).

Twitter has appeared as a primary micro-blogging website, having over 100 million users accomplishing over 500 million tweets each and every day. Twitter has consistently fascinated the users to convey their emotions and perspective about any concerns, brand, company or any other topic of interest with huge audience. Owing to this logic, Twitter is used as an informative source by many organizations, institutions and companies. On Twitter, using only 140 characters users are allowed to share their emotions in the form of tweets. By using slang, acronyms, abbreviations, emotions, short forms can leads to people compacting their statements. People convey their opinions by using sarcasm and polysemy along with this. Hence it is sustain to phrase the Twitter terminology as unorganized. Sentiment analysis is used to extract the sentiments and opinions. The results from this can be used in many fields like analyzing and monitoring changes of sentiment with an event, release of a targeted product and sentiments regarding a particular brand, estimating public view of government policies etc. These type of policies are mainly used by the organizations and also by companies.

II. METHOD

Transfer Learning:

We test their transfer capabilities and generality after choosing the best models and their parameters. We probed whether the ultimate hidden state depiction which can be treated as a projection of the content of the tweet's into a lower dimensional space is convenient only for the task for which it was trained or is it sufficient also for forecasting other emotion categorizations. We take a model up to the ultimate hidden layer and then re-train the final softmax or sigmoid layer on another data set. In this process, we restate the embedding from one data set for making forecasting on the other. Note that since we are copying weights of one model to the other, we are also enforced to use a frequent model architecture; i.e. the number of neurons, layers, type of layers, number of feature maps, size of kernel, etc. The instinct at the rear of these experiments is that if the ultimate hidden state representations can be treated as a generic lower dimensional depiction is applicable for predicting feelings, then the one Ekman might also satisfy for concluding POMS's divisions. Nonetheless, if the trained model performance is drastically worse than that of a initially trained model on POMS, this would indicate that ultimate hidden states representations which are specifically tuned for appropriate categorizations of emotions.

Unison Learning:

The ultimate set of experiments tests whether it is desirable to evolve a common model. We define the unison model as a model which is able to evaluating all the three emotion classifications while sharing all the parameters that project the input tweet into a ultimate hidden state depiction. The utility of such model is at least threefold. First, sharing parameters will hopefully lead to a model whose ultimate hidden state depictions are more generous. The existence of such hidden state that could be used to predict the different emotional classifications is an implication that there endure a general emotion representation, which could be the starting point for evaluating the interdependence between emotional classifications. Second, as it is believed for multitask learning approaches; suggesting these additional signals during the training of a model could lead to better performance. Finally, a single model would require for one classification when applying such model, we get predictions for all classifications in approximately with the same computation time. We propose the following architecture to build the unison model. We have common embedding, followed by a Neural Network layer. There are three different softmax (for multiclass setting) or sigmoid (for multilevel setting) layers, after the ultimate hidden state of representation of the Neural Network, each predicting one of the three classifications. Embedding that is informative enough for predicting all three categorizations at once. A similar idea was presented by Collobert and Weston; however, their tasks were not closely related and consequently they only shared a word embedding. Contradictory, our tasks are such related and hence it makes sense to try sharing the whole RNN or CNN layer as well.

System Model:

To set the standard performance, we first examine with common access to emotion disclosure. Within the dimension of pure machine learning (as averse to using, say emotion lexicons), (BOW) bag-of-words models is a simple classifier which is one of the most generally used method. We decided to use recurrent (RNN) and convolution (CNN) networks, among the most popular neural network (NN). The departed were preferred since they can instinctively handle variable lengths of text, and latter since they have already shown to be suitable for text classification. The tweets were collected and analyze the emotion type in unison model. To one could incorporate morphological variation, synonyms or derivations of words when searching emotional hash tags. Also, the neutral class should be added to properly distinguish between emotional and factual tweets the model can distinguish between emotions based on character set. To study the interdependence between emotion classification and to test the generality of model.

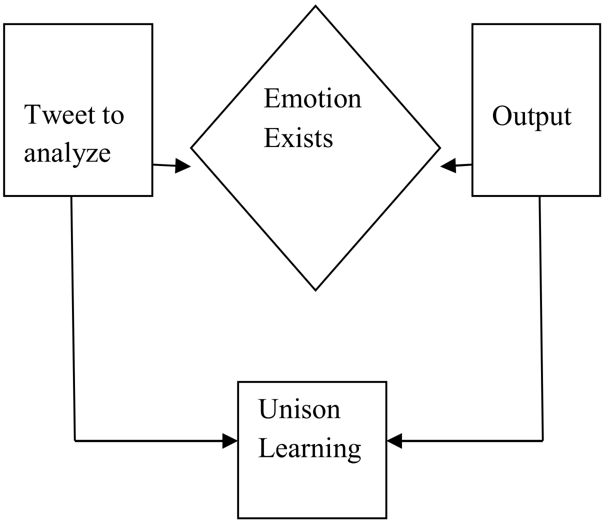

Fig.1-System Model

III. RELATED WORK

Twitter mood to predict the stock markets:

Emotions can profoundly affect individual behavior and decision making tells by behavioral economics. Does this also apply to societies at huge, can societies experience mood states that affect their collective decision making? Is public mood correlated by extension or economic indicators are even predictive? Here we investigate whether measurements of collective mood states obtained from large. The values of the Dow Jones Industrial Average (DJIA) are correlated to twitter feeds over time. By using two mood tracking tools we analyze the content of the text daily from Twitter feed, namely Google Profile of Mood States (GPOMS) that measures mood in terms of 6 dimensions (Calm, Alert, Sure, Vital, Kind, and Happy) and Opinion Finder which it measures positive vs. negative moods. By comparing their capability we cross-validate the resulting mood time series to detect a public's response to the presidential election and Thanksgiving Day in 2008. To examine the hypothesis that are public mood states, a Granger causality analysis and a Self-Organizing Fuzzy Neural Network are used, as measured by the Opinion Finder and GPOMS mood time series, are predictive in changing the closing values of DJIA. Our results can be significantly improved by the inclusion of specific public mood dimensions indicates that the accuracy of DJIA predictions but not others. By predicting the daily up and down changes in closing values of DJIA we find that 87.6% accuracy and the Mean Average Percentage Error of a reduction is more than 6%.

A General Psycho evolutionary Theory of Emotion

The general psycho evolutionary theory of emotion that is emotion that is conferred here that it has many number of extensive characteristics. First, it presents a vast evolutionary framework for conceptualizing the domain of emotion as seen in animals and humans. Second, it can afford a structural model which depicts the correlations among emotions. Third, it can validate both theoretical and factual relations among a number of imitative domains including personality traits, diagnosis, and ego defenses. Fourth, it has contribute the theoretical rational for the development of tests and scales for the measurement of key dimensions within these various domains. Fifth, it has assumed a good pact with factual research using these tools and concepts. Finally, this theory provides helpful institution into the relationships among the emotions, adaptations, and evolution.

Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank

Semantic word spaces have been very convenient but cannot express the meaning of longer phrases in a convention way. Further, the progress towards understanding the compositionality in tasks such as sentiment apprehension requires richer supervised training and evaluation resources and it is more powerful models of composition. To remedy this, we introduce a Sentiment Treebank. It includes the fine grained sentiment labels for 215, 154 phrases in the parse trees of 11,855 sentences and presents new challenges for the sentiment compositionality. We introduce the Recursive Neural Tensor Network to address them. When trained on the new tree bank, this model exceeds all previous methods on several metrics. It pushes the state of an art in single sentence positive or negative classification from 80% up to 85.4%. By predicting the fine grained sentiment labels for all phrases, the accuracy reaches 80.7% an improvement of 9.7% over bag of features baselines. Finally, it is the only model that can accurately capture the effects of negation and its scope at various tree levels for both positive and negative phrases and classify according to the sentiment of the approach in which it is detected easily and also comparatively used by the various communities at large. When we compared to all other model the accuracy will become more and one drawback is it could not find longer phrases.

IV. PROPOSED METHODOLOGY

Twitter is a great platform which is widely used by people to express their opinions and display sentiments on different incidents. To analyze the data and reclaim sentiment by using sentiment analysis that it exemplify. Twitter sentiment analysis is an application of sentiment analysis on data from Twitter (tweets), in order to extract sentiments convey by the user. Sentiment analysis approaches can be predominantly categorized into two classes lexicon based and machine learning based. Lexicon based approach is unsupervised as it proposes to perform analysis using lexicons and a scoring method to check out opinions. Whereas machine learning approach involves the use of feature extraction and training the model using feature set and some dataset.

Advantages of methodology:

- ✓ The dataset collected is imperative for the model efficiency. The deciding factor for the efficiency of the model is the categorization of dataset into training and testing sets.
- ✓ Users opinion to predict the positive, negative and neutral.
- ✓ Easy to predict and determine.
- ✓ Opinion can be classified in an easy aspect.

In order to execute sentiment analysis, we are needed to collect the data from the desired source (here Twitter). This data collected endures various steps of preprocessing which makes it more machine prudent than previous from.

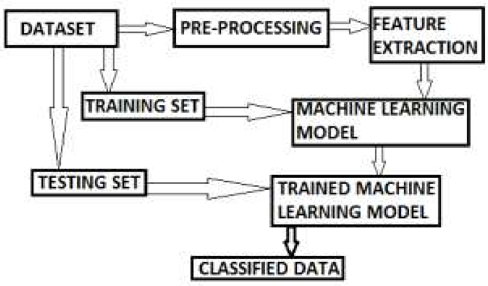

Fig.2 Methodology for sentiment analysis

The proposed work is a little different as they correlate with the event and sentiment by using a timestamp. By using this approach for a distinct event, it is possible to segregate it into sub-events and further expand the study of user sentiments. This approach is convolute but produce very detailed type results when we choose a large event and desire to see variations in user sentiments relating with time.

Natural Language Processing (NLTK)

Natural Language Toolkit (NLTK) is a library in which provides the base for text processing and classification in python. With the use of NLTK, operations such as tokenization, tagging, filtering, manipulation of text can be performed. Various trainable classifiers area also embodies with NLTK library (example Naive Bayes Classifiers). NLTK library is used for creating a bag-of words model, which is a type of unigram model for text. The number of circumstances of each word is counted in this model. The gain of data can be used for training classifier models. By using a sentiment lexicon, the sentiment of the entire tweets is computed by assigning perspicacity score to each word.

V. CONCLUSION

Twitter sentiment analysis comes under the league of text and opinion mining. It mainly targets on investigating the sentiments of tweets and supply the data to a machine learning model so we can train it and the check its accuracy, according to the results we can use this model for future. It contains process like collection of data, pre-processing the text, detection of sentiment, classifying the sentiments, training and testing the model. This research topic has appeared during the last 10 years with models stretching the efficiency of almost 85% to 90%. There is still lacks of data in the dimension of diversity. Along with this it has many of application concern with the slang used and the acronyms of words. When the number of classes are increased many analyzers do not react well. It is not still tested that how accurate the model will be for topics other than one in deliberations. Hence sentiment analysis has a very intense purview of progress in future.

VI. REFERENCES

(1) J. Bollen, H. Mao, and X.-J. Zeng, “Twitter mood predicts the stock market,” J. of Computational Science, vol.2, no.1, pp. 1-8, 2011.

(2) D. Gruhl, R. Guha, R. Kumar, J. Novak, and A. Tomkins, “The Predictive Power of Online

Chatter,"Proc. Of the 11th ACM SIGKDD Int. Conf. on Knowledge discovery in data mining, pp. 78-87, 2005.

(3) G. Mishne and N. Glance, "predicting Movie Sales from Blogger Sentiment," AAAI Spring symposium: Computational Approaches to Analyzing Weblogs, pp. 155-158, 2005.

(4) P. Ekman, "An Argument for Basic Emotions," Cognition & Emotion, vol. 6, no. 3, pp. 169-200, 1992.

(5) R. Plutchik, "A General Psycho evolutionary Theories of Emotion. Academic Press, 1980, vol. 1, pp.3-33.

(6) J. C. Norcross, E. Guadagnoli, and J. O. Prochaska, "Factor structure of the Profile Of Mood States (POMS): Two Partial Replications," J. of Clinical Psychology, vol. 40, no. 5, pp. 1270-1277, 1984.

((7)

(7) S. M. Mohammad and S. Kiritchenko, "Using Hashtags to Capture Fine Emotion Categories from Tweets," Computational Intelligece, vol. 31, no. 2, pp. 301-326, 2015.

(8) S. M. Mohammad, X. Zhu, S. Kiritchenko, and J. Martin, "Sentiment, emotion, purpose, amd style in electoral tweets," Information Processing and Management, vol. 51, no. 4, pp. 480-499, 2015.

(9) C. O. Alm, D. Roth, ad R. Sproat, "Emotions from text: machine learning for text-based emotion prediction," in Proc. Of Human Language Technology Conf. and Conf. on Empirical Methods in Natural Language Processing, no. October. ACL, 2005, pp. 579-586.

(10) B. Plank and D. Hovy, "Personality Traits on Twitter-or-How to Get 1,500 Personality Tests in a Week," in Proc. Of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, 2015, pp. 92-98.

(11) R. Socher, A. Perelygin, J. Y. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts, "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank," in Proc. Of the Conf. on Empirical Methods in Natural Language Processing. Citeseer, 2013, pp. 1631-1642.

(12) O. Irsoy and C. Cardie, "Opinion Mining with Deep Recurrent Neural Networks," in Proc. Of the Conf. on Empirical Methods in Natural Language Processing. ACL, 2014, pp. 720-728.

(13) R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural Language Processing (almost) from Scratch," J. of Machine Learning Research, Vol. 12, pp. 2493-2537, 2011.

(14) A. Radford, R. Jozefowicz, and I. Sutskever, "Learning to Generate Reviews and Discovering Sentiment," arXiv preprint arXiv:1704.01444, 2017.

(15) A. Go, R. Bhayani, and L. Huang, "Twitter Sentiment Classification using Distant Supervision," CS224N Project Report, Stanford, vol. 1, pp. 1-6,2009.

(16) N. Nodarakis, S. Sioutas, A. Tsakalidis, and G. Tzimas, "Using Hadoop for Large Scale Analysis on Twitter: A Technical Report," arXiv preprint arXiv:1602.01248, 2016.

(17) E. Kouloumpis, T. Wilson, and J. Moore, "Twitter Sentiment Analysis: The Good the Bad and the OMG!" Proc. Of the 5th Int. AAAI Conf. on Weblogs and Social Media(ICWSM 11), pp. 538-541, 2011.

(18) R. Gonz`alez-Ib`a~ nez, S. Muresan, and N. Wacholder, "Identifying Sarcasm in Twitter: A Closer Look," in Proc. Of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technonologies, no. 2010. ACL, 2011, pp. 581-586