Twitter emotion detection with spam recognition using machine learning algorithms

PALLAVI DHANVE, POOJA KSHIRSAGAR, SUPRIYA BHOSLE, SIDHDESWARI CHAUDHARI
BE Students, Department of Information Technology, Sinhgad Academy of Engineering.

LAXMAN DEVOKATE
Assist Professor, Department of Information Technology, Sinhgad Academy of Engineering.

Abstract- The online social networks are a very large growth in the world today, but the attacks are more common, including one of the attacks is the attack of Twitter in this spammer spreading several malicious tweets that can take the form of links or hash tags in the website and online services, which are too harmful for real users. To prevent these attacks, training tweets are added and, moreover, these problems are solved by extracting 12 lightweight functions, like the age of the account, no. of followers, no. to follow, no. of tweets, no. of re-tweets, etc. For the transmission of spam detection from tweets, the discretization of a function is important for the performance of spam detection. There is a great truth in the system that includes a total of 600 public tweets based on the URL-based security tool. Spam detection primarily creates the classification model that includes binary classification and can also be solved using the automatic learning algorithm. Machine learning algorithms such as the Naïve Bayesian classifier or the vector support machine classifier have informed the behavior of the models. The system reported the impact of data-related factors, such as the relationship between spam and non-spam, the size of training data and data sampling, and detection performance. The implemented system function is the detection of simple and variable tweets of spam over time. The system shows how spam detection is a major challenge and bridges the gap between performance appraisals and focuses primarily on data, features and patterns to identify the real user and inform the user of spam when providing the valuable response binary. The contribution work is to detect the tweets of emotions in real time, because the new tweets come in the form of sequences and use the updated training data set.

Keywords- Emotion Recognition, Text Mining, Spam, Machine Learning, Twitter.

I. INTRODUCTION

Online social networking sites like Twitter, Facebook, Instagram and some online social networking companies have become extremely popular in recent years. People spend a lot of time in OSN making friends with people they are familiar with or interested in. Twitter, founded in 2006, has become one of the most popular microblogging service sites. Around 200 million users create around 400 million new tweets a day for spam growth. Twitter spam, known as unsolicited tweets containing malicious links that the non-stop victims to external sites containing the spread of malware, spreading malicious links, etc., hit not only more legitimate users, but also the whole platform Consider the example because during the election of the Australian Prime Minister in 2013, a notice confirming that his Twitter account had been hacked. Many of his followers have received direct spam messages containing malicious links. The ability to order useful information is essential for the academic and

industrial world to discover hidden ideas and predict trends on Twitter. However, spam generates a lot of noise on Twitter. To detect spam automatically, researchers applied machine learning algorithms to make spam detection a classification problem. Ordering a tweet broadcast instead of a Twitter user as spam or non-spam is more realistic in the real world.

II. RELATED WORK

In [1] paper, check out whether or not public mood as measured from big-scale series of tweets published on twitter.com is correlated or even predictive of DJIA values. The consequences shows that modifications within the public temper nation can certainly be tracked from the content of large-scale Twitter feeds by way of instead simple textual content processing techniques and that such changes reply to a ramification of socio-cultural drivers in an exceptionally differentiated way. Advantages are: Increases the performance. Public temper evaluation from Twitter feeds gives an automated, fast, unfastened and massive-scale addition to this toolkit that can be optimized to degree a diffusion of dimensions of the public temper nation. Disadvantages are: It avoids geographical and cultural sampling mistakes.

The paper [2] Analyzed financial blogs and on-line news articles to expand a public mood dynamic prediction model for stock markets, referencing the perspectives of behavioral finance and the traits of online economic groups. A public mood time series prediction model is likewise provided, integrating features from social networks and behavioral finance, and uses huge information evaluation to assess emotional content material of commentary on modern inventory or economic issues to forecast changes for Taiwan stock index. Advantages are: It is convenient for feature word expansion and processing speed. More widely used. Disadvantages are: Only uses for stock prices.

In [3] paper the software of deep recurrent neural networks to the challenge of sentence-stage opinion expression extraction. DSEs (direct subjective expressions) consist of specific mentions of personal states or speech events expressing nonpublic states; and ESEs (expressive subjective expressions) encompass expressions that imply sentiment, emotion, etc., without explicitly conveying them. Advantages are: Deep RNNs outperformed previous (semi) CRF baselines; achieving new state-of-the-art results for fine-grained on opinion expression extraction. Disadvantages are: RNNs do not have access to any features other than word vectors.

In [4] paper analyze electoral tweets for extra subtly expressed facts such as sentiment (tremendous or bad), the emotion (pleasure, sadness, anger, and so forth.), the cause or reason behind the tweet (to point out a mistake, to aid, to ridicule, and so forth), and the style of the tweet (simple statement, sarcasm, hyperbole, and many others). There are sections: on annotating textual content for sentiment, emotion, fashion, and categories including cause, and on automatic classifiers for detecting those classes. Advantages are: Using a multitude of custom engineered features like those concerning emoticons, punctuation, elongated words and negation along

with unigrams, bigrams and emotion lexicons features, the SVM classifier achieved a higher accuracy. Automatically classify tweets into eleven categories of emotions. Disadvantages are: Does not summarize tweets. It does not automatically identifying other semantic roles of emotions such as degree, reason, and empathy target.

In [5] paper, i) represent how large amounts of social media data can be used for large-scale open-vocabulary personality detection; ii) evaluate which features are predictive of which personality dimension; and iii) present a novel corpus of 1.2M English tweets (1,500 authors) annotated for gender and MBTI. Advantages are: The personality distinctions, namely INTROVERTEXTROVERT (IE) and THINKINGFEELING (TF), can be predicted from social media data with high reliability. The large-scale, open-vocabulary analysis of user attributes can help improve classification accuracy.

The paper [6] develops a multi-task DNN for learning delegations across multiple tasks, not only leveraging huge amounts of cross-task data, but also benefiting from a regularization effect that leads to more general representations to help tasks in new domains. A multi-task deep neural network for representation learning, in particular focusing on semantic classification (query classification) and semantic information retrieval (ranking for web search) tasks. Demonstrate strong results on query classification and web search. Advantages are: The MT-DNN strongly performs using strong baselines across all web search and query classification tasks. Multitask DNN model successfully combines tasks as disparate as classification and ranking. Disadvantages are: The query classification incorporated either as classification or ranking tasks not comprehensive exploration work.

In [7] article, show that emotion-word hashtags are good manual labels of emotions in tweets. Proposes a method to generate a large lexicon of word emotion associations from this emotion-labeled tweet corpus. This is the first lexicon with real-valued word emotion association scores. Advantages are: Using hashtagged tweets can collect large amounts of labeled data for any emotion that is used as a hashtag by tweeters. The hashtag emotion lexicon is performed significantly better than those that used the manually created WordNet affect lexicon. Automatically detects personality from text. Disadvantages are: This paper works only on given text not synonym of that text.

The paper [8] focuses on studying two fundamental NLP tasks, Discourse Parsing and Sentiment Analysis. The improvement of 3 independent recursive neural nets: for the key sub-obligations of discourse parsing, specifically structure prediction and relation prediction; the 1/3 internet for sentiment prediction. Advantages are: The latent Discourse features can help boost the performance of a neural sentiment analyzer. Pre-training and the individual models are an order of magnitude faster than the Multi-tasking model. Disadvantages are: Difficult predictions to multi-sentential text.

III. SYSTEM OVERVIEW

Emotion Recognition is a psychological instrument for assessing the individual's mood state. It defines 65 adjectives that are rated by the subject on the five-point scale. Each adjective contributes to one

of the six categories. For example, feeling annoyed will positively contribute to the anger category. The higher the score for the adjective, the more it contributes to the overall score for its category, except for relaxed and efficient whose contributions to their respective categories are negative. Mood state combines these ratings into a six-dimensional mood state representation consisting of categories: anger, depression, fatigue, vigour, tension and confusion. Comparing to the original structure, we discarded the adjective blue, since it only rarely corresponds to an emotion and not a color, and word-sense disambiguation tools were unsuccessful at distinguishing between the two meanings. We also removed adjectives relaxed and efficient, which have negative contributions, since the tweets containing them would represent counter-examples for their corresponding category.

Contribution of this paper is to implement the Classifier based approach is given to solve the detection of spam messages. A classification model is mainly based on machine learning algorithm which gives the output in the form of binary value. Here the feature extraction is important phase of project to add more benefits to the system. A performance evaluation is carried out on a large dataset which includes around 600 tweets to identify the spammer also system helps to categories the spam and non-spam message.

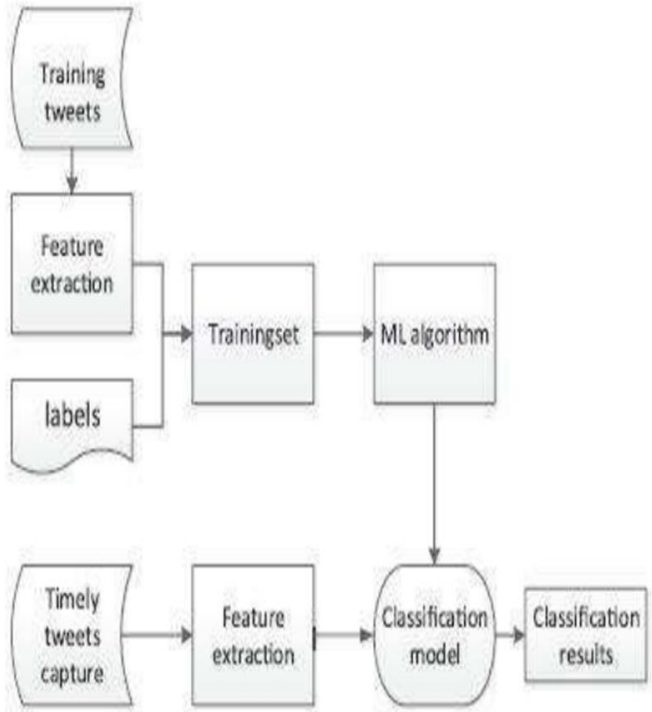

Fig. 1 System Architecture

Advantages are:

- Increases human-computer interactions
- Low-cost
- Fast emotion recognition and Spam detection system
- Scalable
- Comparable quality to experts

A. Algorithms

1. Algorithm: Simplified SMO

Input:

C: regularization parameter

tol: numerical tolerance

max_passes: max # of times to iterate over α's without changing

$(x^{(1)}, y^{(1)}), \ldots, (x^{(m)}, y^{(m)})$: training data

Output:

$\alpha \in R^m$: Lagrange multipliers for solution

$b \in R$: threshold for solution

Process:

1. Initialize $\alpha_i = 0, \forall i, b = 0$.
2. Initialize passes = 0.
3. **while** $(passes < \max\_passes)$
4. num_changed_alphas = 0.
5. **for** $i = 1, \dots, m,$
6. Calculate $E_i = f(x^{(i)}) - y^{(i)}$ using (2).
7. **if** $((y^{(i)}E_i < -tol \;\&\&\; \alpha_i < C) \;||\; (y^{(i)}E_i > tol \;\&\&\; \alpha_i > 0))$
8. Select $j \neq i$ randomly.
9. Calculate $E_j = f(x^{(j)}) - y^{(j)}$ using (2).
10. Save old $\alpha's$: $\alpha_i^{(old)} = \alpha_i, \alpha_j^{(old)} = \alpha_j$
11. Compute L and H by (10) or (11).
12. **if** (L == H)
13. **continue** to next i.
14. Compute η by (14).
15. **if** $(\eta >= 0)$
16. **continue** to next i.
17. Compute and clip new value for α_j using (12) and (15).
18. **if** $(\left|\alpha_j - \alpha_j^{(old)}\right| < 10^{-5})$
19. **continue** to next i.
20. Determine value for α_i using (16).
21. Compute b_1 and b_2 using (17) and (18) respectively.
22. Compute b by (19).
23. num_changed_alphas := num_changed_alphas + 1.
24. **end if**
25. **end for**
26. **if** $(num\_changed\_alphas == 0)$
27. passes := passes + 1
28. **else**
29. passes := 0
30. **end while**

V. RESULT AND DISCUSSIONS

Experiments are done by a personal computer with a configuration: Intel (R) Core (TM) i3-2120 CPU @ 3.30GHz, 4GB memory, Windows 7, MySQL 5.1 backend database and Jdk 1.8. The application is web application used tool for design code in Eclipse and execute on Tomcat server. Some functions used in the algorithm are provided by list of jars like Twitter-core and Twitter-stream jars etc.

Tweets are removed steamingly, and Twitter gives the steaming API for designer and scientists to get to open tweets continuously utilizing Twitter4j containers. The point of this paper is to conquer and hindrance via completing a execution assessment, which was from two distinct angles NLP and machine leaning algorithms. The Unison model is the combination of Ekman's, Plutchik's and POMS emotion categories and the Sequential Minimal Optimization (SMO) classifier algorithm uses for emotion recognition performance. Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. Recall is the ratio of correctly predicted positive observations to the all observations in actual class. F-measure is the weighted average of Precision and Recall. Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observation to the total observations. The Table I shows performance analysis between unison model versus SMO classifier.

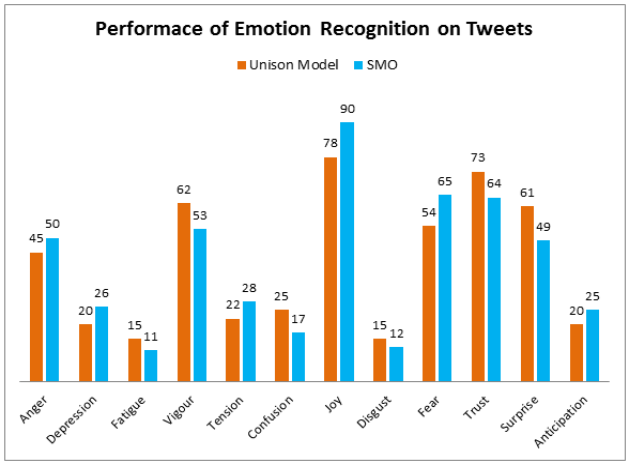

Fig. 2 Comparison of tweets with emotion recognition using Unison Model versus SMO Classifier Algorithm

TABLE I Performance Analysis Between Unison Model Versus SMO Classifier

	Unison Model	SMO
Precision	68.45	78.70
Recall	79.44	65.64
F-Measure	72.11	74.31
Accuracy	80.29	87.26

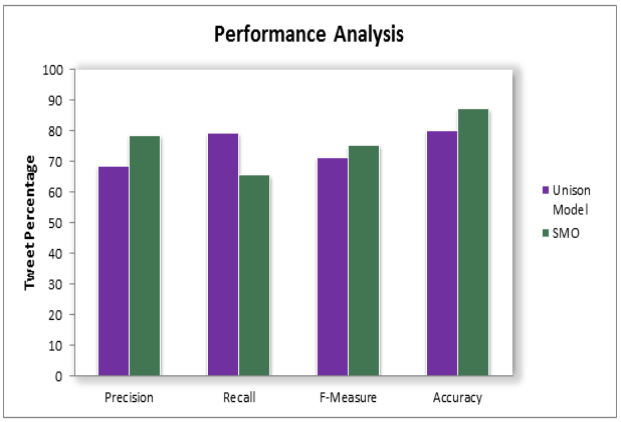

Fig. 3 Performance Analysis between existing and proposed system

Fig. 3 shows the performance analysis between Unison Model and SMO classifier algorithm. The graph shows the Unison Model increases accuracy as compare to previous algorithms. But, the SMO classifier algorithm gives better results than Unison Model. And SMO executes faster than Unison model.

Now, performance analysis on Twitter as well as Facebook users' dataset using emotion recognition. Normally Facebook API is mostly used than Twitter API. So here comparing the emotions detection using both dataset in Fig. 4.

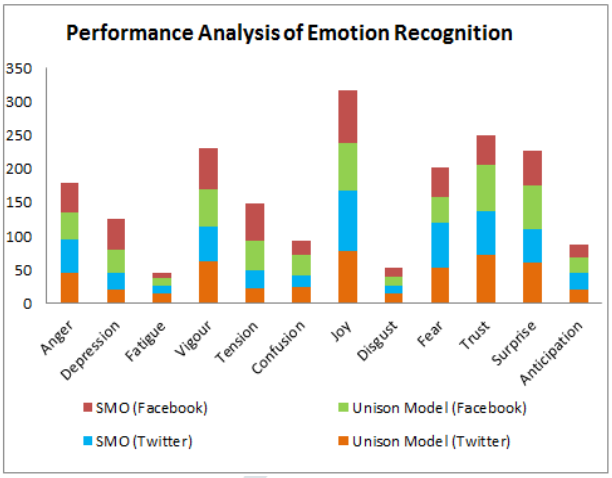

Fig. 4 Performance analysis using Twitter API and Facebook API

IV. CONCLUSION

This project implements a machine learning algorithm represents twelve-dimensional mood state representation using 65 adjectives with combination of Ekman's and Plutchik's emotions categories like, anger, depression, fatigue, vigour, tension, confusion, joy, disgust, fear, trust, surprise and anticipation. These classifies the emotions with the help of bag-of-words algorithm and Classifier based approach is given to solve the detection of spam messages.

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] J. Bollen, H. Mao, and A. Pepe, Modeling Public Mood and Emotion: Twitter Sentiment and Socio-Economic Phenomena, in Proc. of the 5th Int. AAAI Conf. on Weblogs and Social Media Modeling, 2011, pp. 450-453.

[3] S. M. Mohammad, X. Zhu, S. Kiritchenko, and J. Martin, "Sentiment, emotion, purpose, and style in electoral tweets," Information Processing and Management, vol. 51, no. 4, pp. 480–499, 2015.

[4] 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.

[5] X. Liu, J. Gao, X. He, L. Deng, K. Duh, and Y.-Y. Wang, "Representation Learning Using Multi-Task Deep Neural Networks for Semantic Classification and Information Retrieval," Proc. of the 2015 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 912–921, 2015.

[6] 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.

[7] S. M. Mohammad and S. Kiritchenko, "Using Hashtags to Capture Fine Emotion Categories from Tweets," Computational Intelligence, vol. 31, no. 2, pp. 301–326, 2015.

[8] B. Nejat, G. Carenini, and R. Ng, "Exploring Joint Neural Model for Sentence Level Discourse Parsing and Sentiment Analysis," Proc. of the SIGDIAL 2017 Conf., no. August, pp. 289–298, 2017.

[9] S Kamble, SM Sangve,"Real time Detection of Drifted Twitter Spam Based On Features," International Journal of General Science and Engineering Research (IJGSER), ISSN 2455-510X,Vol 4(1), 2018,21-23.

[10] SMS Harshad Dattatray Markad,SM Sangve, "Parallel Outlier Detection for Streamed Data Using Non-Parameterized Approach," IJSE,Volume 8, Issue 2 July-December 2017.

[11] M. Farhoodi and A. Yari, "Applying machine learning algorithms for automatic Persian text classification," 2010 6th International Conference on Advanced Information Management and Service (IMS), Seoul, 2010, pp. 318-323.

[12] E. Tromp and M. Pechenizkiy, Rule-based Emotion Detection on Social Media: Putting Tweets on Plutchik's Wheel, arXiv preprint arXiv:1412.4682, 2014.

[13] S. Chaffar and D. Inkpen, Using a Heterogeneous Dataset for Emotion Analysis in Text, in Canadian Conf. on Artificial Intelligence. Springer, 2011, pp. 6267.

[14] S. Aman and S. Szpakowicz, Identifying Expressions of Emotion in Text, in Int. Conf. on Text, Speech and Dialogue, vol. 4629. Springer, 2007, pp. 196205.

[15] G. Mishne, Experiments with Mood Classification in Blog Posts, in Proc. of ACM SIGIR 2005 Workshop on Stylistic Analysis of Text for Information Access, vol. 19, 2005, pp. 321327.