

INTEGRATION OF MACHINE LEARNING AND SSA OPTIMIZATION FOR PROACTIVE DETECTION OF STRESS

B. Gomathi^{*1}, R.Sujatha^{*2}, S .T.Deepa^{*3},V.Venkateswari^{*4}

¹ Associate Professor, Department of Computer Science, Shri Shankarlal Sundarbai Shasun Jain College For Women T-Nagar, Chennai, Tamilnadu. Email:gomathiphd2k15@gmail.com

² Assistant Professor, Department of Computer Science, PSG College of Arts & Science Coimbatore, Tamil Nādu, Email: sujamca2k5@gmail.com

³ Associate Professor, Department of Computer Science, Shri Shankarlal Sundarbai Shasun Jain College For Women T-Nagar, Chennai, Tamil nadu.
Email:stdeepa@shasuncollege.edu.in

⁴Assistant Professor, Department of Computer Science, Shri Shankarlal Sundarbai Shasun Jain College For Women T-Nagar, Chennai, Tamil nadu,
Email:v.venkateswari@shasuncollege.edu.in

Abstract:

Depression is common nowadays without any age limitations. The primary reason for depression is stress. Early detection of stress and treatment will help to cure the same. There are many approaches to detect stress and this paper focus on sentiment analysis. Detection is done based on details in social media and twitter dataset is used. Sentimental analysis is based on extraction from text. Tensi strength is used and this is basically a system for detection. It detects strength of stress as well as relaxation based on expressions posted in Social media. Both positive and negative strength are captured with scaling to indicate the level of stress. Constraint in Tensistrength is it depends on tweets, and drawback here is dis ambiguity. A single word can have different meanings, for example, word Crane, it implies to bird as well as Machine. In order to have better pre-processing SVM with Ngram is used for better accuracy and precision. Our Proposed method works in two phases. During first phase SVM with Ngram is used to detect stress and in the second phase it is fused with Salp Swarm Algorithm. This fusion provides better result compared with PSO.

Keywords: Stress Detection, SSA algorithm. SVM with Ngram, Sentimental analysis.

1.Introduction:

Nowadays human life is full of deadlines and targets. This target varies with different age group, for school children targets with respect to academic performance, for adolescent age objective is different and especially for working people demands are more and most of them are unable to manage personal and professional life. Human life is tightly packed with deadlines and tensions and this leads to Stress. If the degree of stress is higher this will disrupt life too. Human behaviour can be analysed using sentimental analysis. Tweets, blogs, review are rich in sentimental data sets. Trends nowadays is to study reviews, opinion before buying a product. Most of the companies collect the review and improve the product even before launch of the same^[1]. Opinions are categorized as positive, neutral and negative and emotions,

attitudes plays core part in Sentimental analysis^[2]. Health disorders are due to stress and relaxation is required for happy life^[3]. The post in social media can also help in Stress detection. The likes, dislikes and write up on social media, reaction on a particular problem on social media also identifies stress. Through the Nature of followers, blogs visited and write up on social media, a psychiatrist can find people under pressure, crazy, mad after a typical subject. This also is an indication of stress. Use of social networking sites analyses the state of one's mind and thinking . This paper provides a novel approach for identifying stress and integrating with SSA optimization for better performance.

2.Review Literature

One of the approaches to find whether the stress is short term or long term is to use TensiStrength. Set of constraints are used for prediction. The efficiency here again depends on the nature of tweets ^[5]. Major drawback in twitter is number of characters used in tweet. Finding sarcasm level and ranking neighborhood tweet was utilized for predicting stress^[6]. Manju Venugopalan, et.al, (2015) proposed a model for sentiment classification and introduced domain oriented approach. This approach improved accuracy of classification^[7]. Datamining Classifiers can be used for the purpose of sentimental analysis. The author Anurag P. Jain, et.al, (2015) in his paper compared accuracy of KNN and Random forest and the this novel approach proved KNN outperforms and provides better efficiency for sentimental analysis^[8]. The feature and data set used also plays a vital role in prediction. The author Sida Wang (2012) concentrated on model selection and compared NB,BNB and MNB with SVM [9].Sports are prone to tweets and the author T. S. Zakzouk(2012) performed analysis using machine language with domain as sports news.[10]. Similarly, analysis cannot be restricted to single language and the author A. Abbasi (2008) discussed on how feature selection can be done for web forum.[11]. Hybrid approaches produces better results for Feature selection. Salp swarm algorithm fused with PSO produced better results for feature selection. This novel method improved the efficiency and helped find best features [12]. The major limitation of tweet is size, and Arabic tweets are taken as dataset and different classification approaches are used to extract prominent features^[13]. For finding out the attitude of customers Salp Swarm Algorithm is used for feature selection for Arabic language sentiment analysis by the author Abdulla Issa Alza^[14].He proved SSA out performed grey wolf and Particle swarm.

3.Propsed System

The proposed system work is performed in phased manner. During the first phase, we input the data and perform pre-processing. Unwanted data which is of no use to sentiment classification are removed. In second phase Word Disambiguation such as words with more than one meaning and that carriers different logic based on the position of the words are dealt. In the third phase SVM with ngram is used and it is compared with NB. In the Final stage the results are integrated with Salp swarm algorithm and this fusion provides better results. The below figure 1 describes Architecture for Proactive Stress detection.

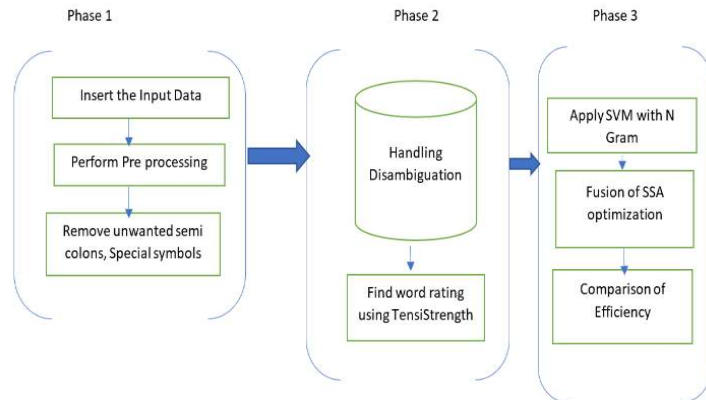


Figure.1 Architecture for Proactive Stress Detection

3.1 Phase 1:

During this stage the first initial step is to upload input data. Generally, this will be text data, this can be from blogs, tweets etc. Pre-processing to be done for the input data given. The most primitive step for sentimental analysis is removing unwanted data like non alphabets, white spaces, special symbols, stop words etc. Apart from this it is normal practice to use shortcuts and this to be converted to full form. References, tags and links to the various websites and URL to be avoided. The below algorithm 1 explains phase 1.

1. Input the Text Data
2. Check whether the data is from blogs, tweets.
3. Remove special symbols, on alphabetic characters
4. Check for null spaces, short cuts, URL
5. Find the start and end of paragraph
6. Pass the output data to the next phase.

3.2 Phase 2:

This phase deals with disambiguous words. There are words that has more than two meanings. The best way to deal with this is Skip gram where only the meaning of the word is taken for rating. Second challenging task is based on the position of the word, the entire scenario changes. Basically, POS tagging which takes only noun, verb, adjectives, adverb is used for processing data. Concepts like Uni, Bi, Tri gram are analysed to remove ambiguous words. Skip gam based on classification of tweets also is used to find stress, relaxed or neutral state. TensiStrength is applied to calculate word rating. Relative scaling +5 to -5 is used where +5 includes happy state or relaxed state and -5 indicates stress or depressed state. Any value within 2.5 to 3.5 indicates neutral state.

3.3 Phase 3

Resultant values of Phase 2 are used in this final stage to detect the current state based on tweets. The results can be either happy state of mind or relaxed, Neutral state and finally depressed or stressed state. Comparison of NB and SVM is done. The output is integrated

with Salp Swarm Optimization algorithm and the results prove that SVM with SSA produced better results. The flowchart for stress deduction is given in the below figure 2.

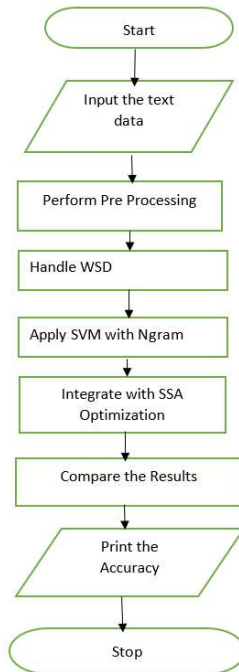


Figure 2 Flow Chart of Proactive Stress Deduction

4. Experimental Analysis

Once word sense ambiguity is clarified the next step is to apply the same to TensiStrength Framework. Classification or prediction of the level of stress that is identifying the index whether it is Happy, neutral or depressed is done. This is accomplished using SVM with Ngram classifier. The performance is compared with Naïve Bayes and SVM with Ngram outperforms NB. The final output are integrated with Salp Swarm optimization and the results prove this fusion yields better results. Data set used is sentiment140, There are around sixteen lakh tweets and the scaling is from +5 to -5.

Attributes of the dataset include

1. Target=Identifying happy, neutral or stressed state.
2. User id: Id of the particular tweet
3. Date: Date of the tweet
4. name: User name of the tweet
5. text: Text involved in tweet
6. Flag: to find the state

A sample population of 200 tweets are taken and categorization is done. The below table 1 shows categorization. Classification is done based on the tweets and apply SVM, NB, SVM SVM_SSA .

Table 1: Categorization Based on tweets

Algorithm	Happy Mind Set	Depressed State	Neutral State	Others
NB	78	32	54	36
SVM_Ngram	76	54	38	32
SVM Ngram_SSA	68	45	33	54

To measure the performance precision and recall is calculated and this is based on True positive value. Based on the confusion matrix the precision, recall and F-Measure is calculated. The below table 2 shows comparison of the different algorithms.

Algorithm	Precision (%)	Recall (%)	Fmeasure(%)
NB	83	68	67
SVM Ngram	74	64	64
SVM Ngram_SSA	87	72	70

Table 2: Performance analysis of Algorithms

The below figure 3 represents comparison of Performance.

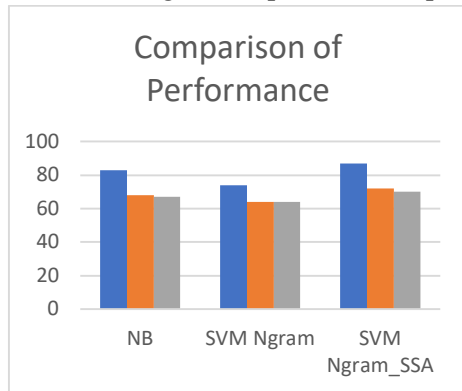


Figure 3: Comparison of Performance

5. Conclusion

Proactive prediction of psychological stress helps human to become better individual. After identification at early stages therapy like yoga and mediation can be suggested based on the level of stress. In the Proposed system based on the tweets, word sense disambiguation is removed and by implementing the ranking in Tensistrength frame classification was made easy. On comparing the algorithms, integration of SSA optimization with SVM yields optimal results. The experimental results prove that fusion of SVM Ngram with Salp Swarm produces

better results than used individually. In future other optimization like PSO, BAT can be implemented with different machine learning algorithms.

References

- [1] Turney, P., “Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews”, 40th annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, Pennsylvania, vol. 19, pp. 134-138, 2002.
- [2] Wilson, T., Wiebe, J., Hoffmann, P., “Recognizing contextual polarity in phrase -level sentiment analysis”, 2012, Human Language Technology and Empirica, vol.23, pp. 672-676
- [3].Reshmi Gopalakrishna Pillai, Mike Thelwall, Constantin Orasan, “Detection of Stress and Relaxation Magnitudes for Tweets”, International World Wide Web Conference Committee ACM, 2018.
- [4]. Munmun De Choudhury, Michael Gamon, Scott Counts, Eric Horvitz, “Predicting Depression via Social Media”, Seventh International AAAI Conference on Weblogs and social media, 2013.
5. Thellwall” TensiStrength: Stress and relaxation magnitude detection for social media texts”, Volume 53, Issue 1, January 2017, Pages 106-121.
6. Prashanth KVTKN,” **A novel method for detecting psychological stress at tweet level using neighborhood tweets**”, 2021, Journal of King Saud University - Computer and Information Sciences.
7. ManjuVenugopalan, Deepa Gupta,” Exploring Sentiment Analysis on Twitter Data”, IEEE Conference Publications, 978-1-4673- 7948-9 vol.6, pp. 34-39, 2015.
- 8.Anurag P. Jain, Mr. Vijay D. Katkar,” Sentiments Analysis Of Twitter Data Using Data Mining”, International Conference on Information Processing (ICIP), 978-1-4673-7758-4, vol. 19, pp. 432-438, 2015
9. Sida Wang and Christopher D. Manning,” Baselines and Bigrams: Simple, Good Sentiment and Topic Classification”, [Proceedings of the 50th Annual Meeting 2012](#)
10. T. S. Zakzouk and H. I. Mathkour, “Comparing text classifiers for sports news,” Procedia Technology, vol. 1, no. 0, pp. 474 – 480, 2012, first World Conference on Innovation and Computer Sciences (INSODE 2011).
11. A. Abbasi, H. Chen, and A. Salem, “Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums,” ACM Trans. Inf. Syst., vol. 26, no. 3, pp. 12:1–12:34, Jun. 2008.
12. REHAB ALI IBRAHIM , AHMED A. EWEES ,DIEGO OLIVA,” IMPROVED SALP SWARM ALGORITHM BASED ON PARTICLE SWARM OPTIMIZATION FOR FEATURE SELECTION”,JOURNAL OF AMBIENT INTELLIGENCE AND HUMANIZED COMPUTING, VOLUME 10, PAGES 3155-3169 (2019)
13. AL-AYYOUB M, JARARWEH Y, RABABAH A, ALDWAIRI M,” FEATURE EXTRACTION AND SELECTION FOR ARABIC TWEETS AUTHORSHIP AUTHENTICATION”, J AMBIENT INTEL HUMANIZ COMPUTE 8(3):383–393,2017
- 14.Abdullah Issa Alzaqebah .”Arabic Sentiment Analysis Based on Salp Swarm Algorithm with S-shaped Transfer Functions, DOI:10.1109/ICICS49469.2020.239507, April 2020