

Survey: Sentiment Stress Identification Using Tensi/Strength Framework

Mrs. Reshma Radheshamjee Baheti, Prof. S.A.Kinariwala

¹MTech student, Department of CSE, MIT Aurangabad, Maharashtra, India

²Professor, Department of CSE MIT Aurangabad, Maharashtra, India

Abstract:

In recent years human stress is most rapidly increasing. The school, collegestudents, job professional and those person who work under pressure. In last few decades researcher's work isgoing on how to predict people under pressure and you must get relax in your duty. In survey we find, if we work on sentiment analysis which will help to find emotions or feelings about daily life. So social media networks like face book, tweeter and other Social networking sites where user can share personal; feeling with friends it may be happy, angry, stress or any type ofemotion which may describe mood of the person. On the social networking sitesa huge number of informal messages, blogs and discussion forums are posted every day. Emotions appear to be frequently vital in these texts forexpressing friendship, presentation social support or as part of online point of view. In this paper we survey on existing techniques which are working to find sentiment analysis of textual data. In textual data must find the positive and negative sentence to decide emotion of user. In survey we find the natural language processing, lexical parser, sentiment analysis, classifier algorithm and some different kinds of tweeter dataset. In our survey we completed85% work on sentiment analysis, here we can easily categories the positive and negative sentence.

Keywords —Stress detection, Sentiment analysis, Data mining, Tensi Strength, natural language processing.

Introduction

Stress is a sentiment of emotional or physical nervousness. It can appear from any occasion or thought that makes you think frustrated, angry, or nervous. Stress is your body's feedback to a challenge or command. In short rupture, stress can be positive, such as when it helps you avoid hazard or meet a target. The stress level are up if people have not getting any kind of satisfaction. It's like different type of work, situation, irritating friends, or partners. In global survey of corporates publish the 6 out of 10 worker are under pressure on duty. In workers, including high professional people, marketing person or which person who might work on deadline.

So we survey on how to predict human stress level. if we search on google "stress" keyword then you getting 1000 link to how to control your stress, then what you does and don't. its help to exercise and diet plan for lunch. But no any other system which can observe your daily life and decide your stress level.A number of different approaches have been utilized, including analyses of morbidity

and mortality by occupational categories. Thus, for example,teachers-professors have mortality rates of arteriosclerotic heart disease that areonly about one half of the rates for physicians-lawyers-pharmacists-insuranceagents. These are jobs of roughly comparable social status, levelof physicalactivity and physical health hazards, and it is not unreasonable to point to workdemands as a possible clue. Within a single occupational group, of physicians,there are differences by specialties that againsuggestively implicate stressfulor demanding work settings; for example, general practitioners show higherrates of mortality and morbidity than specialists. And in a specific organizational setting, such as NASA, prevalence of heart disease was observed to behigher for managers than for scientists and engineers [2].

In survey only 10% people regularlygo to hospital for regular checkup and maintain body for exercise, diet plan and yoga. But remaining people are not going to hospital. Hospital [2] is best solution for checking and predicting stress level and what people do. but it have more expensive to regular checkup.

Common people are not going to hospital and regular checkup.

Now a days People are using social media, like as Twitter and Facebook, to share their feelings and opinions with their acquaintances. Postings on these sites are made in a naturalistic surroundings and in the itinerary of daily actions and goings-on. As such, social media supply a means for incarcerated behavioral characteristic that are relevant to an individual's thoughts, mood, message, activities, and socialization. The feeling and language used in social media postings may indicate feelings of insignificance, responsibility, defenselessness, and self-hearted that characterize major depression. Additionally, depression sufferers often withdraw from social situations and activities. Such changes in movement might be salient with changes in commotion on social media.

So it's an existing problem we must have to implement the unique concept where we analysis of user daily life in specific format. So we survey on existing system which can help me spy on people to daily life to find human stress.

Background and Related work

A. Data mining

Data mining is concept which have a mine the data on different categories pattern format. User have

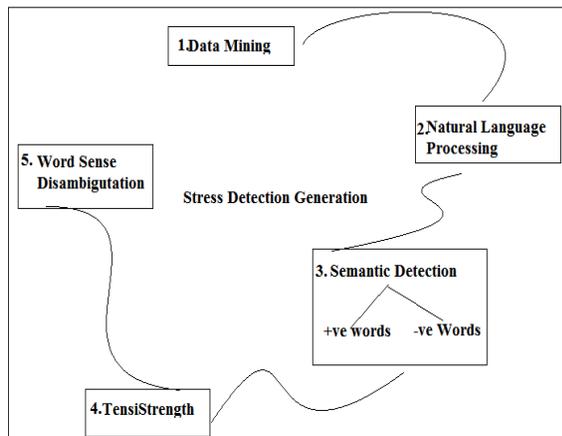


Fig1. Generation of Stress Detection

available raw dataset which contain all types of data related to system. User have to apply any kind of algorithm like prediction based, analysis based, recommendation algorithm and techniques as well as find to similarity [6] of words of datasets. In sentiment analysis have large number of textual

datasets are available which contain the train and test dataset. The train datasets means sorted dataset where compare the test dataset.

B. Natural Language Processing

NLP is techniques used to find sentiment analysis on textual data. The every textual datasets have apply the NLP concept. it contains grammar checking, spelling mistake etc. The NLP has contain the lexical parser [19] classifier is the best for analysis of textual data. Lexical semantics start with a identification that a word is a conventional association between a lexicalized idea and an statement that plays a syntactic role. A lexical matrix, consequently, can be stand for theoretical purposes by a mapping between written words and Syntax. Since English is wealthy in synonyms, synsets are often enough for differential purposes. Synonymy is, a lexical relation between word forms, but because it is assigned this central role in WordNet, a notational difference is made between words related by synonymy, which are with this in curly brace, '{' and '}', and other lexical relations, which will be enclosed in square brackets, '[' and ']'. Semantic relations are pointed by pointers. [17]. a supervised learning method forward sense disambiguation based on Inner Product of Vectors [7]. The lexical move toward is to start with an existing set of terms with known sentiment orientation and then use an algorithm to predict the sentiment of a text based upon the incidence of these words [3].

C. Sentiment Detection

The sentiment analysis can done by using the data mining and NLP techniques. Basically we must find out the emotion of words. Like strength of word [1], [3]. Sentences are processed with NLP processing, it must check the sentence grammar, spelling mistake, remove unwanted data in preprocessing format. to investigate talk about above concerning strength detection for multiple emotions there is some work on positive-negative sentiment strength detection [1]. The existing survey can find out the rating of words for positive and negative base. To calculate average of words its generate the sentiment of words. A common approach for sentiment analysis is to select a machine learning algorithm and a method of extracting *features* from texts and then train the classifier with a human-coded corpus. The features used are typically words but can also be stemmed words or part-of-speech tagged words, and also may be combined into two consecutive words and trigrams. The center of the algorithm is the sentiment word strength catalog. This is a compilation of 298

positive terms and 465 negative terms classified for either positive or negative sentiment strong point with a value from 1 to 5. The default classifications are based upon human opinion during the development stage, with automatic modification occurring later during the preparation phase

1. Positive Words:

Sentiment analysis of words categorizing emotion levels. For every sentence we must find out the rating or scale of words. Positive words means getting the energetic words, if people listen then react on happy, normal, surprise etc. In survey we collect the positive and negative word dictionary. Which contain all word by word ratings.

2. Negative Words:

Negative words are negation in a sentence, not, nor, opposite word of positive words. Its demoralize the people. We had found all negative words which can be used then people will get angry.

D. TensiStrength

TensiStrength is an adaptation of the sentiment strength detection software SentiStrength [9]. The responsibilities of sentiment and stress/relaxation recognition are related but not corresponding. stress seems likely to coincide with high arousal and a negative valence, whereas relaxation may coincide with low arousal and a positive valence [5]. In TensiStrength has getting user out as -5 to 5 rating bases. -5 level is last stage of stress and 5 means its normal level of people. Its contain all the machine learning techniques with different datasets [3]. A text with a score of 3, 5 would contain moderate positive sentiment and strong negative sentiment. A neutral text would be oblique as 1, 1. Two scales are used because even short texts can hold both positivity and negativity and the aim is to detect the sentiment expressed rather than its in general polarity. TensiStrength develop its lexicon by transmission to each sentence to achieve the highest stress term recognized and the highest relaxation term recognized. Multiple sentence texts are allocate the maximum value of any ingredient sentence. [9]. Following are some preprocessing used in TensiStrength

1. Remove Special symbol: In sentiment sentence to start preprocessing we remove the special all symbol.

2. Spelling checker: We can used to social media data, social media content are write in short cut format. We improve the every word spelling.

3. Lingusting : To express emotion textual data write in long format. e.g. coool. we must remove the this kind of data and correct word.

4. Remove Common word: In system to improve performance we must remove the common word of every sentence, like I, am, he she etc. This type of words are not useful but it increases your performance.

4. Word Matching: Getting every word to compare existing train dictionary, where already word have sort out -5 to 5 format. Means stress full word is -5 and most relax word is 5.

5. Sentence Analysis: Every sentence getting multiple of -5 to 5 scale rating. Here you have must analysis and combination of result to increase system accuracy.

The automatic refinement system for term strengths uses a hill-climbing approach by assessing whether altering the score of any term in the lexicon could improve the overall accuracy on the development set. This process randomly selects a term from the TensiStrength lexicon, increments its weighting and then rejects the change unless it improves the sum of the positive and negative scores by at least 2.

Technical Survey

1. Support Vector Machines

The *support vector machine* is a strong inductive learning scheme that enjoys a considerable theoretical and empirical support for using SVMs for text categorization [12]. The SVM computation of the (soft) maximum margin is posed as a quadratic optimization problem that can be solved in time complexity of $O(kn^2)$ where n is the training set size and k is the dimension of each point. Thus, when applying SVM for text categorization of large datasets, an efficient representation of the text can be of major importance [17].

2. TBL classifier

Transformation-Based Learning (TBL) is a general machine learning classification method described comprehensively by Brill (1995). It has been successfully applied to a variety of problems in natural language processing, including POS tagging,

BaseNP identification, text chunking and prepositional phrase attachment. These modifications to the TBL algorithm resulted in statistically significant absolute performance improvements ranging from 1% on both English and Basque data to 2.2% on Spanish data [18].

3. Supervised WSD

WSD is an attractive topic for researchers and is an important method for many NLP relevance such as Information Retrieval, Machine Translation, and speech recognition etc. WSD refers to the procedure of mechanically identifying the correct significance of a confusing word (i.e., a multi-meaning word) based on the context in which it occurs. The supervised advance applied to WSD systems use machine-learning technique from physically created sense-annotated information. Training set will be worn for classifier to learn and this training set consist pattern connected to target word. These tags are manually created from lexicon. It contains some techniques or algorithm of system [17]. A supervised learning method for word sense disambiguation based on Inner Product of Vectors. Using TWA as a benchmark dataset, we first extracted two sets of features; the set of words that have occurred frequently in the text and the set of words surrounding the ambiguous word. Then, using 10 fold cross validation approach the dataset was divided into training and test parts for a context similarity based classifier [7]. The physically additional stress terms and pointer of stressors and stressful condition were derived from a range of academic and non-academic foundation that describe stress in general or stressors connected with travel. Word sense vagueness can be considered as the most serious difficulty in machine exchange systems. The human mind is clever to select the correct target equal of any source language word by sympathetic of the background. a WSD approach that is hold on inner product of vectors algorithm. The proposed scheme is a

supervise approach in which sense-tagged data is worn to train the classifier. WSD approach remove two set of features; the set of words that have Co-occurred with the vague word in the text often, and the set of words nearby the ambiguous word. The main task performed by the disambiguation technique is to allocate a sense to an ambiguous word by measure up to the context it has occurred in and the texts obtainable in the training quantity.

1. Decision List

Its contain a if- else statement.

2. Decision Tree

It's a hierarchy of words where generate a tree. The tree represent the ascending or descending order of system. Internal node of a decision tree denotes a test which is going to be applied on a feature value and each branch denotes an output of the test. When a leaf node is reached, the sense of the word is represented.

Conclusion:

We survey on detecting stress by using social media textual data. In survey we find existing sentiment analysis system where data mining, NLP, lexical parser. and SentiStrength. As well as we study on some classifier like SVM, WSD and TBL Classifiers.

To detecting human stress we implement Support vector machine as well as SentiStrength technique.

Reference

1. Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*.
2. Stanislav V. Kasl, " STRESS AND HEALTH", Ann. Rev. Public Health. 1984. 5:319-41
3. Thelwall, M., Buckley, K., & Paltoglou, G. (2012). Sentiment strength detection for the social Web. *Journal of the American Society for Information Science and Technology*, 63(1), 163-173.
4. Margaret M. Bradley and Peter J. Lang, " Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings", The Center for Research in Psychophysiology, University of Florid.
5. Mike Thelwall, " Sentiment and Stress Analysis with Senti/TensiStrength".
6. L. Douglas Bakerti, Andrew Kachites McCallum, " Distributional Clustering of Words for Text Classification", CM 1-58113-015-5_8/98 \$5.00.
7. Xinxiong Chen, Zhiyuan Liu, Maosong Sun, " A Unified Model for Word Sense Representation and Disambiguation", 2014 Association for Computational Linguistics

8. Reshmi Gopalakrishna Pillai, Mike Thelwall, Constantin Orasan, "Detection of Stress and Relaxation Magnitudes for Tweets", International World Wide Web Conference Committee.
9. Thelwall, M. (in press). TensiStrength: Stress and relaxation magnitude detection for social media texts. *Information Processing & Management*. doi:10.1016/j.ipm.2016.06.009
10. Andrew Trask & Phil Michalak & John Liu, "SENSE2VEC - A FAST AND ACCURATE METHOD FOR WORD SENSE DISAMBIGUATION IN NEURAL WORD EMBEDDINGS.", Under review as a conference paper at ICLR 2016.
11. Eric H. Huang, Richard Socher*, Christopher D. Manning, Andrew Y. Ng, "Improving Word Representations via Global Context and Multiple Word Prototypes", 2012 Association for Computational Linguistics.
12. RON BEKKERMAN, "DISTRIBUTIONAL CLUSTERING OF WORDS FOR TEXT CATEGORIZATION" HAIFA JANUARY, 2003.
13. Rami Al-Rfou, Bryan Perozzi, Steven Skiena, "Polyglot: Distributed Word Representations for Multilingual NLP", 2013 Association for Computational Linguistics.
14. Munmun De Choudhury Michael Gamon Scott Counts Eric Horvitz, "Predicting Depression via Social Media", Association for the Advancement of Artificial Intelligence.
15. Munmun De Choudhury Scott Counts Eric Horvitz, "Predicting Postpartum Changes in Emotion and Behavior via Social Media," 2013 ACM 978-1-4503-1899-0/13/04.
16. Huijie Lin^{1,2}, Jia Jia^{1,2}, Quan Guo³, Yuanyuan Xue¹, Qi Li¹, Jie Huang¹, Lianhong Cai^{1,2}, Ling Feng, "User-Level Psychological Stress Detection from Social Media Using Deep Neural Network", 2014 ACM 978-1-4503-3063-3/14/11
17. Alok Ranjan Pal and Diganta Saha, "WORD SENSE DISAMBIGUATION: A SURVEY", IJCTCM Vol.5, No.3, July 2015.
18. R A D U F L O R I A N, S I L V I U C U C E R Z A N, "Combining classifiers for word sense disambiguation", 2002 Cambridge University Press
19. George A. Miller, Richard Beckwith, Christiane Fellbaum, "Introduction to WordNet: An On-line Lexical Database",