What Makes You Stressed? Finding Reasons From Tweets

Reshmi Gopalakrishna Pillai, Mike Thelwall and Constantin Orasan
Research Institute in Information and Language Processing
University of Wolverhampton, UK
reshmi.g85@gmail.com, {m.thelwall,c.orasan}@wlv.ac.uk

Abstract

Detecting stress from social media gives a non-intrusive and inexpensive alternative to traditional tools such as questionnaires or physiological sensors for monitoring mental state of individuals. This paper introduces a novel framework for finding reasons for stress from tweets, analyzing multiple categories for the first time. Three word-vector based methods are evaluated on collections of tweets about politics or airlines and are found to be more accurate than standard machine learning algorithms.

1 Introduction

Stress is the manifestation of physical or emotional pressure, often as a bodily response to a real or perceived challenge. Selye (1936) defines it as non-specific response of the body to any demand for change. It is an important aspect of the mental state of people including business customers, citizens involved in political debates, and commuters. If detected automatically, it can be used to predict problems such as customer churn, threatening political events or transportation deadlocks in these contexts. In socio-political domains, such as politics, sports, and news, stress detection can help in understanding the stress trends to get a collective mental state of the target population. For example, increases in apparent stress, topics generating stress, or geographical stress hotspots might all have important consequences. Also, for service-centric businesses, including hotels, airports and airlines, in which the owner's goal is to provide a stress-free stay, travel or transit, it is valuable to know the causes of stress for customers, which might point to issues requiring immediate attention.

Social media can be harnessed to discover trends in group or individual emotions and moods. Although previous studies (reviewed below) have developed methods to detect stress in social media, the causes of stress also need to be known so that remedial actions can be targeted more effectively. In response, this research implements a novel framework for finding the causes of stress expressed in tweets. This study introduces a method to classify stress causes from tweets belonging to two domains, one each from socio-political (Politics) and service-centric (Airlines) domains, to demonstrate the viability of the methods.

The contributions of this work are as follows:
1. This is the first multiple category study detecting reasons for stress expressed in tweets.
2. A dataset of tweets annotated with reasons for stress.

2 Related Work

2.1 Stress Detection from Social Media

In recent years, social media content analysis has emerged as a useful tool to evaluate the mental health of users. Internet usage patterns (Kotikalapudi et al, 2012) and status messages on Facebook (Moreno et al, 2011) have been demonstrated to be viable tools for evaluating

266

depressive tendencies. Similarly, message content and interaction patterns on Twitter can also be harnessed to help identify depression (De Choudhury et al, 2013), Post Traumatic Stress Disorder (PTSD) (Coppersmith et al, 2014), and postpartum emotional and behavioral changes (De Choudhury et al., 2013).

TensiStrength (Thelwall et al, 2017) is the first lexical based program to detect the strength of stress and relaxation in tweets. Its lexicon is derived from LIWC (Tausczik and Pennebaker, 2010), General Inquirer (Stone et al, 1986) and emotion terms from the sentiment analysis software SentiStrength (Thelwall et al, 2010; Thelwall et al, 2012). TensiStrength estimates stress (on a scale of -1 to -5) and relaxation (on a scale of +1 to +5) with accuracy comparable to several general machine learning algorithms. The performance of this system was improved by adding word sense disambiguation as a preprocessing step for tweets (Gopalakrishna Pillai et al., 2018).

Though there is a growing interest in finding expressions of stress from social media content, as discussed above, the existing research does not, for the most part, discuss the reasons for stress. Our model, on the other hand, studies the reasons for stress in multiple categories.

2.2   Topic Modelling in Tweets

Topic modelling is the extraction of latent topics in documents, which may be helpful to find stress reasons from a collection of texts. Two common topic modelling methods for documents are Latent Dirichlet Allocation (LDA) (Blei et al, 2003) and Author Topic Models (ATM) (Rozen-Zvi et al., 2005).

The applicability of these methods to tweets is hindered by informal language, grammatical errors, slang and emoticons. To overcome these issues, aggregation of related tweets into individual documents has been proposed as a potential solution, called pooling.

Mehrotra et al (2013) proposed one of the most widely accepted pooling methods to overcome the limited coherence of LDA on Twitter data. It found that pooling tweets by hashtags performs better than other pooling schemes (author-wise, hourly, and burst-wise) based on Point-wise Mutual Information (PMI), NMI scores and purity scores.

Alvares-Melis and Saveski (2016) present a scheme for tweet pooling in which tweets and their replies are aggregated into a single document. The users who participate in the conversation are considered to be co-authors of this pooled document. We used an LDA-based topic modelling with hashtag pooling in our present study. Though conversation pooling was found to give better performance compared to hashtag pooling, it was not suitable for our datasets, which consisted of tweets having the relevant hashtags and could not be grouped into 'conversations'.

2.3   Word Vectors and its Application in Sentiment Analysis

Liu (2012) defines sentiment Analysis is as the field of study that analyses opinions or sentiments of people towards entities such as products, services, individuals and their attributes. Sentiments in text are most often expressed by opinion words which has positive (good, wonderful, fantastic) or negative (bad, poor, horrible) polarity. However, finding the inherent sentiment of a text from content words is not a straightforward problem, due to ambiguity of word meanings and complex sentiments such as sarcasm. Hence, efficient and accurate word representations which considers the context information also, become necessary.

Representation of words as real-valued vectors has been employed in sentiment analysis, as in other NLP problems. There are two common architectures for word vector representations: Word2Vec (Mikolov et al, 2013) and GloVe (Pennington et al, 2014). Word2Vec has two models: Skipgram where the objective is to predict a word's context given the word itself and Bag of Words (BoW) where the objective is to predict a word given its context. GloVe (Global Vectors) was proposed as an alternative model, in which the global corpus statistics are captured directly. Over the years, there have been attempts to incorporate the sentiment information of the words into these vectors, to make them more suitable for analysis of sentiment in documents and short texts such as tweets (Maas et al,2011, Tang et al., 2014). Our methods to find stress reasons from tweets also use word vector representations as illustrated in the next section.

3 Methods

3.1 Overview

The proposed method selects reasons for stress expressed in tweets from a pre-defined list of potential stressors for tweets belonging to two categories, politics and airlines, collected by the Tweepy API. Tweets with high stress scores, as judged by TensiStrength, were considered for creating this list of potential stressors. These high-stress tweets were subjected to topic modelling and k-means clustering to find the clusters of frequently occurring topics. Topic modelling provides a soft clustering of the topics, however we followed it with k-means clustering to obtain coherent collections of topics. These topic clusters were manually refined to generate title words that most aptly encompass each cluster. The title words constituted a list of potential stressors for the tweets of that category.

To automatically detect stress reasons, the tweets were processed by three new word-vector based methods to find a reason for the stress expressed within them. These were compared with reasons found by human coders to evaluate the accuracy.

3.2 Method details

Finding Potential Stressors: The first step is to form a list of potential reasons for stress in a given category/domain.
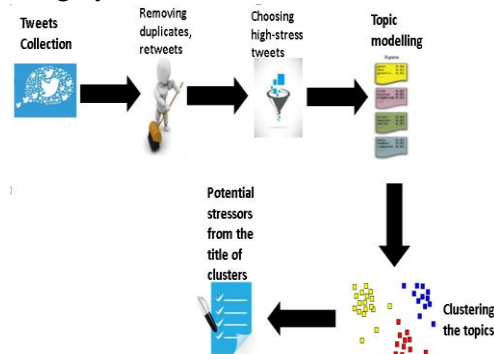

Figure 1: Finding potential stressors for a category/domain.

Word Vector Processing: The tweets were preprocessed to eliminate URLs, prepositions, interjections and conjunctions. Constituent words in hashtags were separated. The remaining words constitute the content words set. Three different word-vector based methods were used to find causes of stress from the list of potential stressors.
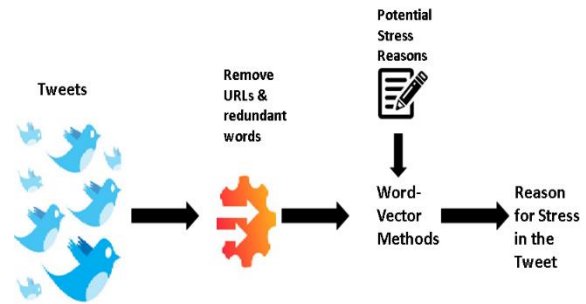

Figure 2: Finding reasons for stress in tweets.

Method 1 (maximum word similarity): The cosine similarity of each word in the content words set was calculated with each potential stressor. The stressor with highest similarity with any of the content words in the tweet was chosen as the stress cause.

Method 2 (context vector similarity): A context vector was found for each tweet by calculating the average of the word vectors of all words in the content words set. The stressor with highest cosine similarity with this context vector was chosen as the stress cause.

Method 3 (cluster vector similarity): Each stressor was represented by a cluster vector which is the average of vectors of all words in its topic cluster. The cosine similarity of each of these cluster vectors was calculated with the context vector and the cluster with maximum similarity was chosen as the stress cause.

3.3 Dataset and Annotation

Two different datasets of public Twitter posts were collected with the Tweepy API.

Politics: For political tweets, the search parameter was the hashtag "#politics AND #us" and #uspolitics from 14.04.2018 to 14.05.2018. This retrieved 22293 tweets, which were processed to remove duplicates, retweets and tweets with only URLs. The resulting dataset had 8163 tweets. The first task was to make a list of potential stressors for tweets which could be used for the further stressor identification tasks. The underlying assumption was that frequently discussed topics in tweets with very high stress scores were potential stressors.

Stress scores were assigned on a scale of -1 (no stress) to -5 (high stress) to each tweet in the dataset, using TensiStrength. The 2205 tweets having a stress score of -5 or -4 were filtered to form the corpus for further processing. They were then preprocessed by removing all URLs,

268

3

@usernames and stop words and divided into groups of 200 tweets each (11 groups, the last one having 205 tweets). The dominant topics in each group were found by an LDA-based topic modelling with hashtag pooling implementation in Python. These topics were aggregated and the k-means clustering algorithm used to separate them into 7 clusters. This number of clusters produced the most coherent and intuitive clusters for this collection.

The seven clusters were manually checked to find the most apt descriptive word for each one, after removing outliers, if any. For example, one cluster had topic modelling key terms: **rape, crime, rage, murder, terrorism, fight, chaos, avalanche, abuse.** We chose to describe this cluster by the word "violence". The title words for all 7 clusters constitute the list of potential stressors. Clusters and potential stressors emerging from them are listed in Table 1.

Example topics in the cluster	Stressor
Vote voter polls candidate race	Election
Public activity support boycott	Protest
Rape crime rage murder terrorism fight chaos avalanche abuse	Violence
Democrats Republican Trump	Person/ Party
Press report news scandal publicity editor	Media
Social system education act government	Policy
Wages inequality employed education college productivity	Economy

Table 1: Clusters and stressors (Politics)

To evaluate the new methods, out of the 8163 tweets obtained after duplicate removal, 4517 tweets with expressions of stress were selected (TensiStrength scores, -5, -4 or -3). 2000 tweets were randomly chosen from this collection and were annotated individually and independently by three human coders. Their task was to select the most appropriate stressor from the predefined list of potential stressors produced by the topic modeling. Coding guidelines were provided and inter-coder agreement scores were calculated using Krippendorff's α (Krippendorff, 2004) and Pearson's correlation. The values, given in Table 2, were high enough to justify the use of the human codes.

Agreement Between	Krippendorff's α
A and B	72.54
B and C	75.95
A and B	73.17

Table 2: Inter-coder agreement for stressor annotation (Politics)

Airlines: A similar process was followed to create the Airlines dataset. The tweets were obtained by searching for hashtags belonging to 9 popular airlines (#gojetairlines, #allnipponairways, #airnewzealand, #swissair, #turkishairlines, #airfrance, #unitedairlines, #emirateairlines, #ryanair), during the same period as the political tweets. The search returned 31457 tweets and, after duplicates and retweets removal, 7965 tweets. Out of this, 3214 tweets were found to have expressions of high stress, (stress score -5 or -4) using TensiStrength system. These were analyzed by topic modelling to find out the list of potential stressors in the category, as detailed in the previous section.

The 3214 tweets having stress values of -5 or -4 were divided into groups of 300 (11 groups, the last group having 214 tweets) and using topic modelling with hashtag pooling we found out the topics in each groups; which was aggregated and further analyzed by k-means clustering to form five clusters after manual refining to remove the outliers. Examples of topics in the five detected clusters and the stressor title word corresponding to each of them are given in Table 2.

Example topics in the cluster	Stressor
Cost money ticket airline expensive	Cost
Delay delayed hours time today cancellation	Delay
Service customer staff food pilot	Service
Strike messed hijack attack	Violence
Luggage issue carry missing stolen	Luggage

Table 3: Clusters and stressors (Airlines)

Out of the 7965 tweets after duplicate removal, 4367 had stress scores of -3 or above, and we chose 2000 tweets from this randomly, to be annotated for stress reasons. The inter-coder agreement between the three coders is given below in Table 4.

Agreement Between	Krippendorff's α
A and B	71.23
B and C	76.19
A and B	78.23

Table 4: Inter-coder agreement for stressor annotation (Airlines)

High inter-coder agreement values in both categories denote that the problem definition and guidelines are well-defined and followed.

3.3 Experimental Setup

For training the word vectors used in the experiments, a Twitter Word2Vec model trained on 400 million tweets was used, released as part of an ACL W-NUT tasks (Godin et al, 2015).

We ran three machine learning algorithms as comparison baselines.

• AdaBoost: An adaptive boosting algorithm based on a simple classifier.

• Logistic Regression: Simple logistic regression.

• SVM: Support Vector Machines using sequential minimal optimization.

The classifiers were implemented using their default configurations in Weka 3.6. Term unigrams, bigrams and trigrams and their frequencies were the features used. Punctuation was included as a term, with consecutive punctuation treated as a single term (e.g., emoticons, multiple exclamation marks). Cross-sentence bigrams and trigrams were not allowed.

This feature selection was adapted from a similar task of finding the stress and relaxation magnitudes of tweets, in our previous research work TensiStrength (Thelwall, 2017).

4 Results

4.1 Results Summary

The stress reasons were found using the three methods discussed in the previous section. Based on Pearson correlations and exact match percentages with the human annotated scores, the cluster vector method best detects stress reasons (Tables 5, 6).

Method	Accuracy
max. word	47.81
context vector	54.63
cluster vector	**63.41**
SVM	52.48
AdaBoost	50.64
Logistic	49.23

Table 4: Performance of stress reason detection methods in Politics tweets

Method	Accuracy
max. word	50.13
context vector	59.74
cluster vector	**67.29**
SVM	58.13
AdaBoost	54.85
Logistic	52.15

Table 5: Performance of stress reason detection methods in Airlines tweets

4.2 Distribution of reasons

The percentage of tweets with different reasons of stress, according to the cluster vector method, are given in figures 3 and 4.
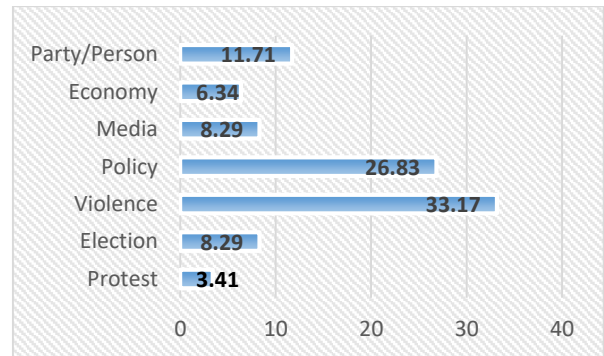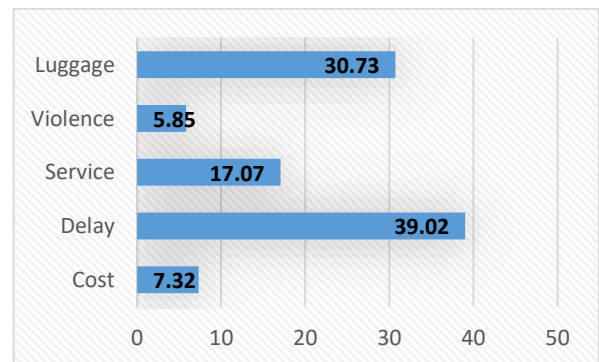

Figure 3: Stress reasons (%) – Politics.

Figure 4: Stress reasons (%) – Airlines.

It is unsurprising that in many (34%) political tweets, violence is the cause of stress and in 39% of airline tweets, delay is the reason. This can be applied in identifying areas of urgent improvement in customer centric businesses.

4.3 Error analysis

There are some systematic reasons for the methods failing to find the correct stress reason.

Misleading hashtags or content words: "At least 14 killed in hockey team's bus crash #news #CNN" This tweet has hashtags #news #CNN which makes all word-vector based methods choose the reason as "media" instead of "violence". "Stocks dive amid fears of trade war" is another example. The human annotated stress reason is "economy" but, war is a misleading word which causes method 1 to choose "violence" as stressor. In methods 2 and 3 where the aggregated tweet vector instead of vectors of individual words are considered, the stressor economy is correctly identified.

Multiple stressors: Tweets in which there are multiple reasons for stress. E.g.: "Killing opponents is a ruthless way to win in elections" has two stressors, "election" and "violence". Expanding the methods to accommodate multiple stressors (by choosing all stressors with cosine similarity with the tweets/content words in tweets above a threshold) will improve its performance in such tweets.

5 Conclusion and Future Work

This paper described three new methods for finding reasons for stress in Tweets list. Datasets comprising of 2000 tweets for Politics and Airlines were manually annotated for stress reasons. The methods found stress reasons more accurately than standard machine learning, although it had problems when multiple causes were expressed in the same tweet, or when key words in the tweet were misleading.

This is the first multi-category study on finding stress reasons in tweets, though limited by the restriction to two domains (politics and airlines) and one source (Twitter). Future work needs to analyze the other domains and also automate the method to detect the potential stress reasons for different domains.

References

David Alvarez-Melis and Martin Saveski. 2016. Topic Modeling in Twitter: Aggregating Tweets by Conversations. ICWSM.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. J. Mach. Learn. Res. 3 (March 2003), 993-1022.

Glen A. Coppersmith, Craig T. Harman, and Mark Dredze. 2014. Measuring post-traumatic stress disorder in Twitter. In Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM).

Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting depression via social media. In Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM).

Munmun De Choudhury, Scott Counts, and Eric Horvitz. 2013. Predicting postpartum changes in emotion and behavior via social media. In Proceedings of the ACM Annual Conference on Human Factors in Computing Systems (CHI), 3267–3276

Sally S. Dickerson and Margaret E. Kemeny. 2004 Acute Stressors and Cortisol Responses: A Theoretical Integration and Synthesis of Laboratory Research. Psychological Bulletin, 130, 355-391. http://dx.doi.org/10.1037/0033-2909.130.3.355

Frederic Godin, Baptist Vandersmissen, Wesley De Neve and Rik Van de Walle. 2015. Multimedia Lab@ ACL W-NUT NER shared task: Named entity recognition for Twitter microposts using distributed word representations. In: Proceedings of ACL-IJCNLP (p. 146).

Reshmi Gopalakrishna Pillai, Mike Thelwall, Constantin Orasan. 2018. Detection of Stress and Relaxation Magnitudes for Tweets. In The 2018 Web Conference Companion (WWW'18 Companion), April 23-27, 2018, Lyon, France, ACM, New York, NY, 8 pages. DOI: https://doi.org/10.1145/3184558.3191627

Elias Jónsson and Jake Stolee. "An Evaluation of Topic Modelling Techniques for Twitter."

Raghavendra Kotikalapudi, Sriram Chellappan, Frances Montgomery. 2012. Associating Internet usage with depressive behavior among college students. IEEE Technol SocMag.;31:73–80.

Klaus Krippendorff. 2004. Content analysis: An introduction to its methodology. Thousand Oaks, CA: Sage

Bing Liu. 2012. Sentiment Analysis and Opinion s Mining. Morgan & Claypool Publishers.

271

Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. 2011. Learning word vectors for sentiment analysis. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1 (HLT '11), Vol. 1. Association for Computational Linguistics, Stroudsburg, PA, USA, 142-150.

Rishabh Mehrotra, Scott Sanner, Wray Buntine, and Lexing Xie. 2013. Improving LDA topic models for microblogs via tweet pooling and automatic labeling. In Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval (SIGIR '13). ACM, New York, NY, USA, 889-892. DOI=http://dx.doi.org/10.1145/2484028.2484166

Tomas Mikolov, Chen Kai, Greg Corrado, and Jeffrey Dean. 2013 Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. 2013. "Distributed representations of words and phrases and their compositionality." In Advances in neural information processing systems, pp. 3111-3119.

Megan A. Moreno, Lauren A. Jelenchick, Katie G. Egan, Elizabeth Cox, Henry Young, Kerry E. Gannon, Tara Becker. 2011. Feeling bad on Facebook: Depression disclosure by college students on a social networking site. Depression and Anxiety, 28(6), 447–455.

Jeffrey Pennington, Richard Socher, and Christopher Manning. "Glove: Global vectors for word representation." In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pp. 1532-1543. 2014.

Michal Rosen-Zvi, Thomas Griffiths, Mark Steyvers, and Padhraic Smyth. 2004. The author-topic model for authors and documents. In Proceedings of the 20th conference on Uncertainty in artificial intelligence (UAI '04). AUAI Press, Arlington, Virginia, United States, 487-494.

Hans Selye. 1956. The Stress of Life. New York, McGraw-Hill Book Company, Inc.

Philip J. Stone, Dexter C. Dunphy, Marshall S. Smith and Daniel M. Ogilvie. 1966. The general inquirer: A computer approach to content analysis. Cambridge, MA: The MIT Press.

Duyu Tang, Furu Wei, Nan Yang, Ming Zhou, Ting Liu, Bing Qin. (2014). Learning Sentiment-Specific Word Embedding for Twitter Sentiment Classification. 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014 - Proceedings of the Conference. 1. 1555-1565. 10.3115/v1/P14-1146.

Yla R. Tausczik and James W. Pennebaker, 2010. The psychological meaning of words: LIWC and computerized text analysis methods. Journal of language and social psychology, 29(1), 24-54.

Mike Thelwall. 2017. TensiStrength: stress and relaxation magnitude detection for social media texts. Journal of Information Processing and Management. 53: 106–121

Mike Thelwall, Kevan Buckley Georgios Paltoglou, D. Cai and A. Kappas. 2010. Sentiment strength detection in short informal text. Journal of the American Society for Information Science and Technology, 61(12), 2544–2558.

Mike Thelwall, Kevan Buckley, and Georgios Paltoglou. 2012. Sentiment strength detection for the social web. J. Am. Soc. Inf. Sci. Technol. 63, 1 (January 2012), 163-173. DOI=http://dx.doi.org/10.1002/asi.21662

Liang-Chih Yu, Jin Wang, K. Robert Lai, and Xuejie Zhang. 2018. Refining Word Embeddings Using Intensity Scores for Sentiment Analysis. IEEE/ACM Trans. Audio, Speech and Lang. Proc. 26, 3 (March 2018), 671-681. DOI: https://doi.org/10.1109/TASLP.2017.2788182

272