Detection of Stress and Relaxation Magnitudes for Tweets

Reshmi Gopalakrishna Pillai
Research Institute in Information and
Language Processing
University of Wolverhampton
Wolverhampton, UK
reshmi.g85@gmail.com

Mike Thelwall
Research Institute in Information and
Language Processing
University of Wolverhampton
Wolverhampton, UK
M.Thelwall@wlv.ac.uk

Constantin Orasan
Research Institute in Information and
Language Processing
University of Wolverhampton
Wolverhampton, UK
C.Orasan@wlv.ac.uk

ABSTRACT

The ability to automatically detect human stress and relaxation is crucial for timely diagnosing stress-related diseases, ensuring customer satisfaction in services and managing human-centric applications such as traffic management. Traditional methods employ stress-measuring scales or physiological monitoring which may be intrusive and inconvenient. Instead, the ubiquitous nature of the social media can be leveraged to identify stress and relaxation, since many people habitually share their recent life experiences through social networking sites. This paper introduces an improved method to detect expressions of stress and relaxation in social media content. It uses word sense disambiguation by word sense vectors to improve the performance of the first and only lexicon-based stress/relaxation detection algorithm TensiStrength. Experimental results show that incorporating word sense disambiguation substantially improves the performance of the original TensiStrength. It performs better than state-of-the-art machine learning methods too in terms of Pearson correlation and percentage of exact matches. We also propose a novel framework for identifying the causal agents of stress and relaxation in tweets as future work.

CCS CONCEPTS

• **Information Systems** → Information Retrieval → Retrieval tasks and goals → Sentiment Analysis • **Human Centered Computing** → Collaborative and Social Computing → Social Media • **Computing Methodologies** → Artificial Intelligence → Natural Language Processing

KEYWORDS

Stress; social media; Twitter; sentiment analysis; word sense disambiguation; word vectors

This paper is published under the Creative Commons Attribution 4.0 International (CC-BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.
WWW 2018, April 23-27, 2018, Lyon, France.
© 2018 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC BY 4.0 License.
ACM ISBN 978-1-4503-5640-4/18/04
DOI <https://doi.org/10.1145/3184558.3191627>

ACM Reference format:

R. Gopalakrishna Pillai, M. Thelwall, C. 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>

1 INTRODUCTION

The ability to identify stress and relaxation is important in the health domain as a diagnostic tool for health-related disorders, for customer focused business to detect unpleasant experiences (e.g., tourism), and for systems that manage humans on a large scale (e.g., transportation, crowd management). The early detection of stress can help action to be taken to avoid the situation escalating or going untreated. Depending on the context, the actions might include recommending a medical consultation for a stress-related disorder, changing hotel operating procedures to avoid customer stress hotspots, re-routing stressed customers to human agents rather than automatic phone agents, or triggering emergency traffic management measures. Similarly, it is important to measure and know the causes of relaxation, as an affective state opposite to stress and as an indicator of satisfaction with applications and services.

Traditionally, stress and relaxation have been measured by monitors identifying variations in physiological parameters or questionnaires targeted at examining the onset of psychological manifestations. These methods, though widely accepted, have limitations. Sensors measuring physical parameters can be expensive and time consuming to use. On the other hand, diagnosis of stress and relaxation through psychological questionnaires is often influenced by manipulated answers given by the participants who might give the responses to project a better image of themselves rather than being honest about their mental state.

Social Networking Sites are popular for both personal and broadcast communication. People habitually share updates from their life on social media platforms. Their online activities and interactions can be expected to mirror the offline incidents and states to some extent. This can be harnessed to evaluate the stress and relaxation levels of a user in a non-intrusive manner.

The detection of stress and relaxation from social media content is a largely unexplored research field. The current study

implements a vector-based word sense disambiguation (WSD) system to improve the performance of the first ever lexicon-based stress/relaxation detection method TensiStrength [1] on Twitter. Results show that incorporating WSD as a preprocessing phase significantly improves the accuracy of TensiStrength and this modified system performs better than standard machine learning algorithms as well. The novel contributions are summarized as follows:

1. Incorporates WSD with stress/relaxation detection, for the first time in social media content research.
2. Analyzes stress/relaxation expressions in tweets belonging to a variety of domains such as traffic, politics, life events and sports.
3. Introduces an updated lexicon with stress/relaxation scores for affect words which accommodates polysemy.

2 LITERATURE REVIEW

2.1 Traditional Methods of Stress/Relaxation measurement

The seminal work, ‘The Stress of Life’ [2] defines stress as the non-specific response of the body to any demand for change. This definition puts forward, two distinct but related aspects of stress: The stimulus i.e., the demand for change and the response. The latter can manifest itself in different forms, mainly as variation in physiological parameters. Stress has been traditionally measured by monitoring indicative parameters such as heart rates [3], galvanic skin response [4] and pupil diameter [5]. Another method is to use questionnaires to find stress-prone Type A personalities [6] or stress-inducing life events [7].

These methods are mostly reactive in nature, requiring the researchers to continuously monitor expensive sensors or rely on users who might be reluctant to share direct observations on their own mental states or manipulate their responses to meet self-imposed psychological images. These limitations provide motivation to develop novel, unobtrusive sources of information to analyze stress and relaxation expressions.

2.2 Identification of Stress and Relaxation from Social Media Content

The ubiquity of social media makes it a potential tool for behavioral and mental health evaluation. Analysis of internet usage pattern [8] and Facebook status updates [9] have been established as unconventional, innovative methods to identify people at risk of depression. There have been efforts to study the content of Twitter, for indicators of mental-health disorders such as depression [10] or Post Traumatic Stress Disorder (PTSD) [11].

The potential for mining social media postings for indicators of mental illness is covered in [12] including the multimedia opportunities and ethical challenges. In a related work, tweets are considered as inputs for predictive models about the postpartum emotional and behavioral changes in new mothers [13]. Using observations on prenatal behaviors, and specifically

attributes such as engagement, emotion, ego network and linguistic styles, these models could classify mothers who would exhibit significant postpartum changes with an accuracy of 71% and 80-83% while additionally considering the initial postnatal data as well.

Similarly, in [10], crowd-sourced Twitter user data is utilized to build a classifier that predicts risk estimation before the onset. This study further analyzes the dataset characteristics to reach observations on diurnal patterns, social connectedness and volume of postings for both depression and non-depression classes. The model was able to predict the onset of depression with an accuracy of 72.384% with all features and reduced dimensions.

[11] uses an automated analysis followed by manual refinement to find positive and negative samples for PTSD. The tweets posted by these users were collected during a time window to create a corpus of positive and negative PTSD data, to train three different classifiers. These were used to identify and evaluate PTSD trends in tweets across military and civilian populations. This method was further enhanced [14] for indicators of bipolar disorder, major depressive disorder and seasonal affective disorder.

TensiStrength is the first published system to detect the strength of stress and relaxation expressed in tweets. It is primarily a lexical approach. The lexicon is partly derived from LIWC [15], General Inquirer [16] and emotion terms from SentiStrength [17] a similar lexicon-based sentiment analysis program. For the original evaluation, a corpus of 3066 English tweets were human-coded for stress and relaxation on five point scales: -1 denoting no stress and -5 denoting the highest stress; +1 for no relaxation and +5 for the highest relaxation. Its performance was similar to several machine learning algorithms, including Support Vector Machines.

The hybrid system presented in [18] combines a factor graph model and convolutional neural network (CNN) and estimates the relationship between users’ psychological stress levels and their social network interactions. This method improves stress detection by 6-9% in terms of F1-score and was used to investigate the social interactions of stressed and non-stressed users.

Moodee[19] a practical mobile application addresses challenges of stress detection from social media content, including missing data, time-series modelling and data sparsity problems. It extracts the tweet-level linguistic, visual and social attributes defined in [20]. These attributes are fed into cross auto-encoders which are embedded in a CNN, integrating them to user level content attributes [21]. This system recommends links to users to dissipate stress.

The system described in [22] identifies stress in social media texts. This uses stressor event categories and stressor subjects and a collection of words related to each category and subject on the basis of word embeddings. The stressor event and subjects are identified using a novel hybrid model which combines multi-task learning with CNN. Tweets are assigned a stress value based on the Social Readjustment Rating Scale [7]. The stress

scores obtained this way give results comparable to some state of the art machine learning models.

As reviewed above, research so far primarily focuses on the health implications of stress, trying to identify and remedy the symptoms in a reactive manner. There has been very little analysis of the underlying causes of stress and relaxation. In contrast, one pioneering attempt to find the causes of stress and relaxation [7], limits its area of interest to personal events such as divorce, death and childbirth. Our study, on the other hand, analyzes tweets expressing short-term or long-term stress and relaxation in a variety of contexts such as traffic, politics, personal events and academia.

2.3 Word Sense Disambiguation in Sentiment Analysis

An important challenge for detecting expressions of stress or any affective state is that social media text uses non-standard grammar and informal language. Affective words can be ambiguous, with their sense changing according to the context. For example, in “There was a chill in the air”, the affect word “chill” indicates stress whereas in “I am a pretty chill guy”, it indicates relaxation. A system to resolve the meaning of this word is essential to correctly identify stress. The natural language processing task of WSD identifying the meaning of words in context in a computational manner [23], is the traditional response to this issue.

A WSD system [24] using Babelfy and SentiWordNet illustrates how it can improve the accuracy of sentiment detection in Twitter and SMS test data. Babelfy is a multilingual graph-based method for disambiguating word senses. It is based on the semantic network BabelNet 3.0. The tweets were preprocessed using the tokenizer of Carnegie Mellon University (CMU) Twitter NLP tool and NLTK. The tweets were represented as vectors with three features – sums of corresponding positive, negative and neutral scores of words in SentiWordNet. A supervised Random Forest Decision Trees classifier was trained based on these representations. This system was shown to have better accuracy (58.55%) compared to the baseline method (45.26%). The paper establishes that WSD improves sentiment analysis accuracy of social media data, specifically tweets.

A research work on sentiment analysis for figurative language [25] applies WSD and assigns polarity to word sense through a graph-based method. Further, sentence-level polarity is detected based on two Hidden Markov Models each trained for positive and negative cases. This system has significantly better recall and precision compared to a polarity detection method without WSD and a baseline WSD method (which assigns the first sense entry in WordNet to all senses).

Incorporating WSD based on the context of word-of-mouth documents to modify SentiWordNet lexicons [26] is found to improve sentiment analysis performance. Similarly, a WSD system [27] uses a path based semantic relatedness to find the most appropriate sense and it is found to give a better classification f-score for sentiment analysis task.

Yet another study [28] integrates a WSD algorithm using extended gloss overlap described in [29] with VoxPop [30], a

classifier of text polarity based on SentiWordNet and improved accuracy from 50.5% to 60.0%.

While varying in methodology and lexical resources, these studies illustrate that sentiment analysis systems with WSD substantially outperform those without WSD. However, the implications of incorporating WSD in stress/relaxation analysis have never been studied before and our research treats it as a potential way to improve the accuracy of an existing lexical method.

2.4 Word Vector Representation and its Application in Word Sense Disambiguation

Standard natural language processing methods for WSD are predicated on standard grammar and spelling, which is uncommon in the social web. A promising method for WSD for informal text is to represent a word as a real-valued vector registering the frequency of its contextual words (e.g., in the same sentence). This can be compared to similar aggregate vectors for the different senses of the word and the closest matching vector used to select the most likely word sense. GloVe, an unsupervised learning algorithm [30] and Word2Vec [31, 32] are the two most common architectures for finding vector representations for words. Word2Vec uses continuous vector representations of words in two models – Bag of words and Skip-gram. In bag of words, the target is predicting a central word in focus, given the input context words. The Skip-gram model, on the other hand, has the focus word as the single input vector and the target is to predict the context words. The GloVe model aims to combine the benefits of the skip-gram model with global statistical information about word co-occurrences using matrix factorization methods.

Sense2Vec [34] is a supervised sense embedding system which can distinguish between different senses of the same word based on the context. The system leverages supervised NLP labels to determine the sense in an instance. This reduces the computational overhead making it more suitable for computationally intensive NLP tasks. It is demonstrated as an efficient WSD method in the context of traditional NLP tasks such as Named Entity Resolution, Sentiment Disambiguation and Part of Speech tagging.

Similarly, a sense vector scheme obtained from skip-gram based word vectors and WordNet glosses has been proposed for WSD [34]. The algorithm presented in this paper works as two steps: initialization of word and sense vectors and sense disambiguation. In the first step, the word vectors are learned through skip-gram model and from this, sense vectors for each sense is constructed. For each word in the gloss, the cosine similarity with the original word is calculated. Those words with cosine similarity higher than a threshold are added to a candidate set. The average of the vectors of the candidate words is taken as the sense vector.

In the second step of word disambiguation, given a sentence, an initial context vector is calculated by finding the average of content words’ vectors. The sense which has the highest cosine similarity to the context vector is taken as the disambiguation result. We implemented this method of sense disambiguation with word vectors in our current experiments.

We chose this unified vector model or word sense representation and disambiguation in our study for the following reasons:

1. It uses a predefined inventory of senses. In our implementation, we pre-computed these vector representations, prior to actual tweets' processing, thus minimizing the computational overhead.
2. This system comprehensively covers all senses in the standard lexical resource, WordNet, representing each by a corresponding vector. It makes use of the high-quality glosses in WordNet, thus forming the basis for word senses with reliable semantic contexts.

3 METHODOLOGY

3.1 Overview

The purpose of this research is to enhance the existing TensiStrength method to identify stress/relaxation strengths by incorporating a pre-processing phase to disambiguate ambiguous affect words in tweets. This is performed by implementing the WSD solution proposed in [34], by finding out which sense vector, among a set of candidate senses, has the highest cosine similarity to the vector representing the context/sentence.

The performance of the improved version of TensiStrength with an incorporated WSD system is compared to that of classic TensiStrength and is found to be consistently better for stress/relaxation accuracy.

3.2 Workflow

The basic approach, devised in the original TensiStrength system and followed in this modified version is to assign to each sentence two scores: those of the highest stress term and the highest relaxation term found in it, modifying these scores with a number of rules covering issues like spelling, negation and booster words. The dual scoring is based on the observation that the same sentence can possibly contain expressions of both stress and relaxation.

The rules considered while assigning stress/relaxation values in original and modified versions of TensiStrength are the following:

1. Two or more repeated letters increase the stress/relaxation values by 1. (scarryy has a higher stress value than scary.)
2. Idioms are treated as single unit with assigned stress/relaxation value. The scores of individual words are ignored.
3. Negation of stress words neutralises them.
4. Negation of relaxing words turn them as stress words.
5. Emoticons are assigned appropriate stress/relaxation values.
6. Spelling correction to delete repeated letters to form words.

Figure 1: Workflow of tweets' processing by the modified TensiStrength algorithm (with WSD)

In the current work, we modified TensiStrength, by adding a pre-processing step, which implements WSD (Figure 1). It first selects a sense out of a list of senses from WordNet using Word2Vec. This pre-processed tweet with the correctly disambiguated affect word is then given to the lexicon based stress/relaxation magnitude detection method with an extended lexicon to include different scores for different word senses rather than a single score for all word senses, as in the previous TensiStrength.

Table 1. Examples of tweets with stress/relaxation values assigned by TensiStrength with and without WSD (with affect (stress/relaxation) terms highlighted)

Tweet	TensiStrength without WSD	TensiStrength with WSD
The future looked uncertain and glum	(-3, +1)	(-3, +1)
Everyone in the valley has a heavy heart but the best thing to do was worship	(-3, +2)	(-3, +2)
She was wearing a cool pink dress	(-1, +2)	(-1, +2)
The lady gave me cool stare	(-1, +2)	(-2, +1)

In the first example (Table 1), there are two stress terms, uncertain and glum with stress strengths -3 and -2 respectively. The score of the highest stress term is assigned to the tweet. In the second tweet, there are two affect words, one each for stress and relaxation and the tweet is assigned stress and relaxation scores corresponding to these words.

TensiStrength without WSD assigns a relaxation strength of +2 and stress strength -1 to the last example. In the modified version, the WSD module identifies the right sense of the affect word 'cold' in the given context and assigns stress strength of -2 and relaxation strength of +1.

3.3 Extending lexicon of TensiStrength to incorporate ambiguity of affect words

TensiStrength uses a lexical approach with a list of terms annotated with strengths of stress and relaxation. TensiStrength’s lexicon is a combination of terms from LIWC [15], General Inquirer [16] and the lexicon of the sentiment detection program, SentiStrength [17] with a few manual additions. We identified 40 ambiguous affect words present in the existing TensiStrength lexicon and manually updated it to accommodate the stress/relaxation values for different senses of each of them. It is to be noted that while the language contains several more ambiguous words, we have, in this experiment, considered only those ambiguous words already present in the TensiStrength lexicon as affect words. A complete list is given in Table 2.

For example, the lexicon for TensiStrength without WSD had only one entry for the affect word ‘cool’, as +2 indicating moderate relaxation. But the modified lexicon, acknowledges the various senses of the same word, obtained from the resource WordNet. For example, the ninth sense of the word ‘cool’ in WordNet is ‘unfriendly or unresponsive or showing dislike’ which has an indication for stress, hence it is assigned the value of -2. Similarly, different senses of each of these ambiguous words are assigned appropriate stress/relaxation values too. The ordinal number of the sense is represented by a suffix to the original word. A few example additions to the lexicon, together with the assigned stress/relaxation strength and the WordNet gloss is given in the table.

Table 2. List of the ambiguous affect words whose stress/relaxation scores were updated in the modified TensiStrength lexicon.

fine	jam	fair	nervous	block
concern	dark	tight	fire	breakdown
chill	wreck	heavy	strain	Fume
Tax	cold	cool	choke	Stick
late	turbulence	fraught	radical	Noise
waste	critical	trick	protest	Stress
scene	mad	hard	scheme	Accident
artificial	brazen	callous	coarse	Desert

Table 3. Examples of additions made to the TensiStrength lexicon

	Stress/relaxation scores	Sense description from WordNet
fine_1	-3	fine, mulct, amercement (money extracted as a penalty)
fine_2	-3	issue a ticket or a fine to as a penalty
fine_3	2	all right, fine, o.k., ok, okay, hunkydory, cool (being satisfactory or in satisfactory condition)
cool_2	3	aplomb, assuredness, cool, poise, sangfroid
cool_9	-2	psychologically cool and unenthusiastic; unfriendly or unresponsive or showing dislike

3.4 Dataset and Annotation

A set of one thousand tweets with ambiguous affect words (e.g. fine) was collected over the period of one month (from 1st February 2017 to 1st March 2017) as the dataset for the experiments. It was annotated individually and independently by a set of three human coders on a five point scale for stress and relaxation.

The coders had participated in an identical annotation procedure as part of the TensiStrength experiment. Hence they had prior experience with the stress/relaxation strength annotation task. The tweets were annotated for stress and relaxation strengths with scores ranging from +1 to +5 (+1 denotes no relaxation and +5 very high relaxation) and -1 to -5 (-1 denotes no stress and -5 very high stress).

The arithmetic mean of the three coders’ annotation scores were calculated for each tweet and rounded off to the nearest integer value. This was taken as the gold standard of stress and relaxation values. The inter-coder agreement between each pair of coders was calculated using Krippendorff’s α [37] and was found to have high positive values. Krippendorff suggests lowest conceivable limit of $\alpha > 0.667$. The overall agreement between the three coders for stress (0.778) and relaxation (0.781) was found to be high enough. It is compared with agreement between the coders in similar experiments in the following table.

Table 4. Inter-coder agreement for stress value annotation

Method	A and B	B and C	A and C
Krippendorff’s α	0.774	0.781	0.771
Pearson Correlation	0.796	0.814	0.792

Table 5. Inter-coder agreement for relaxation annotation

Method	A and B	B and C	A and C
Krippendorff’s α	0.751	0.802	0.781
Pearson Correlation	0.796	0.820	0.797

Higher agreement values in the subsequent experiments denote that the coders are consistently getting better with the annotation task. It also indicates that the problem statement and annotation instructions are well-defined.

3.5 Experimental Setup

A Twitter Word2Vec model trained on 400 million tweets, released as part of an ACL W-NUT task [19] is used for training the word and sense vectors for this experiment. For each word in the ambiguous words’ list mentioned in the previous section, we calculate vectors of dimension 200, corresponding to each different sense, treating WordNet as the sense inventory. The WSD module based on the method presented in [11] was implemented using Anaconda Python (versions, Anaconda-4.4.0 and Python-3.6.1) packages, specifically Scipy, Numpy and nltk, as part of our research work to evaluate the improvement it would bring about in stress and relaxation strength detection.

4 RESULTS

4.1 Performance of TensiStrength with WSD

The accuracy of stress and relaxation detection in tweets using TensiStrength with WSD is compared with TensiStrength without the WSD preprocessing phase and a range of standard machine learning algorithms: Adaptive Boosting algorithm(AdaBoost), Naïve Bayes classifier(Bayes), Decision tree(J48 Tree), Logistic Regression(Logistic) and Support Vector Machines(SVM). Unigrams, Bigrams and Trigrams were used as features. Each classifier was implemented using its configuration in Weka 3.6. The performance of the machine learning algorithms was evaluated using 10-fold cross validation 30 times, with the average scores across the 30 iterations recorded.

Based on Pearson's correlations and exact match percentages with the human-annotated data, it is found TensiStrength with a preprocessing phase to disambiguate the word senses performs considerably better than the state-of-the-art machine learning algorithms and TensiStrength without the preprocessing. The results are summarized in the tables 5,6.

Table 5. Stress detection results

Method	Pearson Correlation	Exact match %	Match %(within 1)
AdaBoost	0.3543	30.1671	83.1562
Bayes	0.3238	35.3464	90.1824
J48 Tree	0.4649	47.4562	88.3543
Logistic	0.4948	49.2679	89.2167
SVM	0.5124	51.2543	91.5626
TensiStrength Without WSD	0.4735	48.8112	83.2367
TensiStrength with WSD	0.5443	53.1091	92.4137

Table 6. Relaxation detection results

Method	Pearson Correlation	Exact Percentage	Match percentage (within 1)
AdaBoost	0.3254	34.1335	83.5451
Bayes	0.3577	38.9221	84.6735
J48 Tree	0.5224	51.5634	86.1734
Logistic	0.4987	54.3267	89.1798
SVM	0.5546	58.7324	91.6598
TensiStrength Without WSD	0.5304	56.3878	85.8364
TensiStrength with WSD	0.56441	60.6981	93.1227

It can be observed that the WSD phase substantially improves the stress/relaxation detection accuracy in terms of Pearson's correlation, exact match percentage and match percentage within 1. TensiStrength with WSD outperforms the machine learning methods as well. However, it should be noted that the small size of the annotated dataset (and the training data) could have been a major cause of the relatively poor performance of

the machine learning methods. Further results with higher number of tweets could better establish the superior effectiveness of TensiStrength with WSD over standard machine learning methods as a stress and relaxation detection. Also, tweets in the current dataset contain the ambiguous words added to the TensiStrength lexicon. The performance on a randomly chosen set of tweets could be different.

4.2 Error Analysis

The misinterpreted stress/relaxation expressions occur in tweets with no direct affect words or affect words with context words misleading the WSD module. In the first category, we can find tweets such as 'I am a train wreck' or 'I had a heavy heart to carry'. The correct (human annotated) stress score for 'I am a train wreck' is -3 and for 'I had heavy heart to carry' it is -4. However, the WSD module incorrectly disambiguates the sense of the ambiguous affect word (wreck, heavy) because the context words train and carry, and assigns a lower stress score of -2 to both. TensiStrength without WSD, gives the scores solely dependent on the lexicon scores of the affect words present.

Table 7. Examples of errors

Tweet	Corr ect score	TensiStre ngth Score (without WSD)	TensiStreng th Score (WSD)	Affect Word	Context Word
I am a human train wreck	(-3, +1)	(-3, +1)	(-2, +1)	Wreck	train
I had a heavy heart to carry	(-4, +1)	(-2, +1)	(-2, +1)	heavy	carry
Great night mad crowd, thanks everyone	(-1, +3)	(-1, +2)	(-3, +1)	mad	crowd
Antioxidant s eliminate destructive potential of free radicals	(-1, +2)	(-2, +1)	(-2, +1)	radical	destructive

Another source of error is indirect expressions of stress and relaxation. Examples in our dataset are 'Sat by the fire; little things in life' (correct stress relaxation score is (-1, +4), our system score is (-2, +1)). 'Straining ears for a laughter clip; but hearing nothing' (correct score (-2, +1) our system score (-1, +3)). Such indirect expressions are challenging for all automated systems in finding the sentiment or stress strengths.

4.3 Domain Analysis of Stress/relaxation for tweets

We manually classified the tweets into one of these domains: Academics, Business, Climate, Entertainment, Food, Health,

Personal Events, Politics, Religion, Sports and Travel. The analysis of stress/relaxation values and performance of TensiStrength with and without WSD are summarized in the following tables, Table 8 and Table 9.

Table 8. Stress detection performance of TensiStrength with and without WSD in different domains

Domain	% of tweets	Mean Stress value	Pearson correlation for TensiStrength (Without WSD)	Pearson correlation for TensiStrength (With WSD)
Academics	10%	-3	0.356	0.412
Business	7%	-3	0.422	0.455
Climate	11%	-3	0.485	0.512
Entertainment	8%	-2	0.487	0.523
Food	6%	-1	0.365	0.376
Health	9%	-3	0.418	0.452
Personal	18%	-3	0.501	0.532
Politics	11%	-4	0.567	0.603
Religion	5%	-4	0.574	0.631
Sports	7%	-3	0.452	0.597
Travel	8%	-3	0.423	0.482

Table 9. Relaxation detection performance of TensiStrength with and without WSD in different domains

Domain	% of tweets	Mean Relax value	Pearson correlation for TensiStrength (Without WSD)	Pearson correlation for TensiStrength With WSD)
Academics	7%	1	0.378	0.425
Business	7%	1	0.523	0.527
Climate	11%	2	0.514	0.497
Entertainment	8%	2	0.390	0.426
Food	6%	2	0.590	0.653
Health	9%	2	0.623	0.671
Personal	18%	2	0.612	0.634
Politics	14%	2	0.633	0.645
Religion	5%	1	0.542	0.602
Sports	7%	2	0.552	0.574
Travel	8%	2	0.587	0.694

1. The largest categories of tweets are personal events, politics and climate categories. The rest of the tweets are almost uniformly distributed across the remaining 8 domains. It can be seen that WSD consistently improves the performance of TensiStrength algorithm in tweets belonging to all domains. The performance improvement is not restricted to any single domain.

2. The tweets have higher values of stress compared to relaxation. The mean value of stress expressed across all domains is -2.81 whereas the mean value of relaxation expressed is +1.72. This could be due to the fact that stress is an affective state with high arousal and relaxation has low arousal. In that sense, stress is a stronger affective state and thus more probable in eliciting responses like tweets.
3. Politics and religion has higher stress values for the tweets (-4) and food and entertainment are domains with lowest stress values (-1 and -2 respectively). This is in agreement with the intuitive expectation of stress values from the domains.

5 CONCLUSIONS AND FUTURE WORK

In this research work, we implemented a WSD solution as a pre-processing stage to an existing lexicon-based stress/relaxation method. A dataset comprising of 1000 tweets with ambiguous affect words was collected and annotated with high inter-annotator agreement. Incorporating a WSD was found to significantly improve the performance of TensiStrength in terms of Pearson's correlation and exact match percentage, for both stress and relaxation. TensiStrength with WSD outperforms machine learning methods as well. Given the relatively small size of the test set, this has to be further studied using bigger datasets annotated with stress and relaxation strengths.

Our planned future step is to analyze the causes of stress and relaxation (namely, stressors and relaxers) expressed in tweets. In the literature review, this was found to be a largely unexplored research field. We formulated a novel framework for finding the stressors and relaxers in a variety of domains such as politics, sports, traffic and personal events. This framework will primarily consist of two modules: Categorizer and Cause Finder. We plan to use a machine learning classifier in the first module to tag the tweets with a broad category of topic such as politics, sports, entertainment, personal events, health, climate and traffic, as listed in the results section. Each category has a list of associated potential stressors and relaxers, to be manually compiled as part of the research. Thus, the reason for expressed stress and relaxation is narrowed down to one among this list. The Cause Finder module finds the actual cause from this list using word vector representations. It eliminates redundant words such as prepositions, conjunctions, interjections and articles from the tweet and constructs a set of keywords from the remaining words. The cosine similarity of vectors representing each word in this set with each potential stressor/relaxer in the relevant category. The stressor/relaxer word represented by the vector having the highest cosine similarity to any of the keywords in the tweet is chosen as the cause of stress/relaxation in the tweet. We ran pilot experiments with a small dataset of tweets to verify the feasibility of this approach and are currently working on compiling the complete dataset and running further experiments.

Thus this research work contributes to the automatic detection of stress and relaxation by both improving the accuracy and

putting forward a novel framework for identifying its causal agents. In future, this can potentially replace or complement traditional stress/relaxation detection methods and can contribute to improved applications and services such as healthcare, traffic management or customer care. It is a promising step in towards the greater goal of leveraging the Web and specifically social media data.

ACKNOWLEDGEMENTS

This research was supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No 636160-2, the Optimum project www.optimumproject.eu.

REFERENCES

- [1] Mike Thelwall. 2017. TensiStrength: stress and relaxation magnitude detection for social media texts. *Journal of Information Processing and Management*. (2017); 53: 106–121
- [2] Hans Selye. 1956. *The Stress of Life*. New York, McGraw-Hill Book Company, Inc.
- [3] J. Choi and R. Gutierrez-Osuna. Using heart rate monitors to detect mental stress, *Proc. 6th Int. Workshop Wearable Implantable BSN*, pp. 219–223, 2009.
- [4] Y. Shi, N. Ruiz, R. Tailb, E. Choi and F. Chen. 2007. "Galvanic skin response (GSR) as an index of cognitive load", *Proc. CHI Ext. Abstr. Human Factors Comput. Syst.*, pp. 2651–2656.
- [5] P. Ren, A. Barreto, J. Huang, Y. Gao, F.R. Ortega, and M. Adjouadi. 2014. Off-line and on-line stress detection through processing of the pupil diameter signal *Ann. Biomed. Eng.*, 42 (1) pp. 162–176, [10.1007/s10439-013-0880-9](https://doi.org/10.1007/s10439-013-0880-9)
- [6] F. Kittel, M. Kornitzer and M. Dramaix. 1986. Evaluation of type A personality. *Postgraduate Medical Journal* 62(730):781–3.
- [7] T.H. Holmes and R. H. Rahe. 1967. The social readjustment rating scale. *Journal of psychosomatic research*, Elsevier
- [8] R. Kotikalapudi, S. Chellappan, F. Montgomery. 2012. Associating Internet usage with depressive behavior among college students. *IEEE Technol Soc Mag.*, 31:73–80.
- [9] M. A. Moreno, L. A. Jelenchick, K. G. Egan, E. Cox, H. Young, K. E. Gannon et. al. 2011. Feeling bad on Facebook: Depression disclosure by college students on a social networking site. *Depression and Anxiety*, 28(6), 447–455.
- [10] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz. 2013. Predicting depression via social media. In *Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM)*.
- [11] G. A. Coppersmith, C. T. Harman, and M. Dredze. 2014. Measuring post traumatic stress disorder in Twitter. In *Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM)*.
- [12] M. De Choudhury, S. Counts and E. Horvitz. 2013. Social media as a measurement tool of depression in populations. In *Proceedings of the 5th Annual ACM Web Science Conference (WebSci '13)*. ACM, New York, NY, USA, 47–56. DOI=<http://dx.doi.org/10.1145/2464464>.
- [13] M. De Choudhury, S. Counts, and E. 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.
- [14] G. Coppersmith, M. Dredze, C. Harman. 2014. Quantifying Mental Health Signals in Twitter. In: *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*. ACL, 2014, p. 51–60.
- [15] Y. R. Tausczik and J. W. Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology*, 29(1), 24–54.
- [16] P. J. Stone, D. C. Dunphy, M. S. Smith and D. M. Ogilvie. 1966. *The general inquirer: A computer approach to content analysis*. Cambridge, MA: The MIT Press.
- [17] M. Thelwall, K. Buckley G. 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.
- [18] H. Lin, J. Jia, J. J. Qiu, et al. 2017. Detecting stress based on social interactions in social networks. *IEEE Transactions on Knowledge and Data Engineering*.
- [19] H. Lin, J. Jia, J. Huang, et al. 2016. Moodex: An Intelligent Mobile Companion for Sensing Your Stress from Your Social Media Postings. *Thirtieth AAAI Conference on Artificial Intelligence*.
- [20] H. Lin, J. Jia, J. Huang, et al. 2014. Psychological stress detection from cross-media microblog data using deep sparse neural network. In *proceedings of IEEE International Conference on Multimedia & Expo*.
- [21] H. Lin, J. Jia, J. Huang, et al. 2014. User-level psychological stress detection from social media using deep neural network. In *Proceedings of ACM Int. Conference on Multimedia*.
- [22] H. Lin, J. Jia, L. Nie, G. Shen, and T. S. Chua. 2016. What does social media say about your stress? In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, 3775–3781.
- [23] R. Navigli. 2009. Word Sense Disambiguation: A survey. *ACM Computing Surveys*, 41 (2009), 1–69.
- [24] C. Sumanth and D. Inkpen. 2015. How much does word sense disambiguation help in sentiment analysis of micropost data? In *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 115–121, Lisbon.
- [25] V. Rentoumi, G. Giannakopoulos, V. Karkaletsis, and G. Vouros. 2009. Sentiment analysis of figurative language using a word sense disambiguation approach. In *Proc. of the International Conference RANLP'09*, pages 370–375, Borovets, Bulgaria.
- [26] C. Hung, S. J. CHEN, 2013: Word sense disambiguation based sentiment lexicons for sentiment classification *Decision Support Systems*, vol. 55, no. 3, pp. 685–697.
- [27] I. Hulpus, N. Prangnawarat, C. Hayes. 2015. Path-based semantic relatedness on linked data and its use to word and entity disambiguation, *Proc. 14th Int. Semantic Web Conf.*, pp. 442–457.
- [28] B. Razon and C. Cheng. 2011. Word Sense Disambiguation of Opinionated Words Using Extended Gloss Overlap. 2011. In *Proceedings of the 8th National Natural Language Processing Research Symposium*, 1–5, Manilla.
- [29] S. Banerjee and T. Peterson. 2003. Extended Gloss Overlaps as a Measure of Semantic Relatedness. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, 802–810, Mexico.
- [30] G. Z. Bautista, M. A. Garcia and R. J. Tan. 2010. VoxPop: Automated Opinion Detection and Classification with Data Clustering. Manila.
- [31] J. Pennington, R. Socher and C. D. Manning. 2014. Glove: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (October 2014), 1532–1543.
- [32] T. Mikolov, K. Chen, G. Corrado and J. Dean. 2013. Efficient Estimation of Word Representation in Vector Space. *arXiv preprint arXiv:1301.3781*, 1–12.
- [33] T. Mikolov, I. Sutskever, K. Chen, G. Corrado and J. Dean. 2013. Distributed Representations of Words, *Advances in neural information processing systems*, 3111–3119.
- [34] A. Trask, P. Michalak, and J. Liu. 2015. Sense2vec-a fast and accurate method for word sense disambiguation in neural word embeddings. *arXiv preprint arXiv:1511.06388*.
- [35] X. Chen, Z. Liu and M. Sun. 2014. A unified model for word sense representation and disambiguation. In *EMNLP*. 1025–1035.
- [36] F. Godin, B. Vandersmissen, W. De Neve and R. Van de Walle. 2015. Named entity recognition for Twitter microposts using distributed word representations. In *Proceedings of the Workshop on Noisy User-generated Text*, pages 146–153, Beijing, China, July 2015. Association for Computational Linguistics.
- [37] K. Krippendorff. 2004. *Content analysis: An introduction to its methodology*. Thousand Oaks, CA: Sage.