

CrystalFeel at SemEval-2018 Task 1: Understanding and Detecting Emotion Intensity using Affective Lexicons

Raj Kumar Gupta and Yinping Yang*

Institute of High Performance Computing (IHPC)

Agency for Science, Technology and Research (A*STAR), Singapore

{*gupta-rk, yangyp*}@ihpc.a-star.edu.sg

Abstract

While sentiment and emotion analysis has received a considerable amount of research attention, the notion of understanding and detecting the intensity of emotions is relatively less explored. This paper describes a system developed for predicting emotion intensity in tweets. Given a Twitter message, CrystalFeel uses features derived from parts-of-speech, n-grams, word embedding, and multiple affective lexicons including Opinion Lexicon, SentiStrength, AFFIN, NRC Emotion & Hash Emotion, and our in-house developed EI Lexicons to predict the degree of the intensity associated with fear, anger, sadness, and joy in the tweet. We found that including the affective lexicons-based features allowed the system to obtain strong prediction performance, while revealing interesting emotion word-level and message-level associations. On gold test data, CrystalFeel obtained Pearson correlations of .717 on average emotion intensity and of .816 on sentiment intensity.

1 Introduction

While humans experience emotions every day, the degree of one's emotions varies from one experience to another. To date, a vast majority of NLP and computational linguistics research deals with ground truth data constructed through the assignment of discrete labels to text messages by annotators. Conventionally, sentiment analysis seeks to determine the valence (positive, negative or neutral) of the feelings and opinions that annotators can recognize in a text message (Hu and Liu, 2004; Pang and Lee, 2008; Socher et al., 2013). Emotion classification, a closely related task, typically seeks to predict the presence or absence of an emotion, i.e., if there is joy or no joy, anger or no

anger, fear or no fear, in a particular message (Alm et al., 2005; Aman and Szpakowicz, 2007; Wen and Wan, 2014). The detection of emotion intensity along a continuous scale is a relatively less explored task.

One of the key reasons for the lack of work on detecting emotion intensity is plausibly attributable to the difficulty in measuring the very concept of emotion intensity. As highlighted in prior research, the question “how intense was your emotional experience on a scale of 1 to 10?” cannot generate reliable responses even for the same emotion type (Frijda et al., 1992). For example, asking people to respond to “how intense was your fear towards getting rejected” and “how intense was your fear towards receiving a medical test result” would lead to inconsistent answers across the same annotators at different times, as well as across different annotators. Because of the lack of a clear reference point, it is nearly impossible to construct ground truth datasets with adequate reliability.

To address the measurement issue, Mohammad and Bravo-Marquez (2017) used a best-worst scaling (BWS) method to create a tweet emotion intensity dataset. Annotators were asked to rank the best and worst examples of the intensity of emotions among n text examples (called n -tube, where $n > 2$ and typically $n = 4$). This reduces the reference point ambiguity issue faced by annotators with regards to which baseline they would have used to rate a text along a single scale. Upon having a target tweet annotated with 24 ranking judgements, the emotion intensity score for the tweet was computed as a real-valued score in the range of 0 to 1 (based on linear transformed value of the difference between the percentage of the number of times the tweet ranked the highest and the times ranked the lowest among all ranking judgements). In total, the dataset consists of 7,097 an-

*Both authors contributed to this research equally. Correspondence should be sent to yangyp@ihpc.a-star.edu.sg.

	Emotion intensity score in range of [0, 0.5]	Emotion intensity score in range of (0.5, 1.0]
Anger	No words Sir... Thank you for the concern.. [0.197]	Everything I order online just comes looking like a piece of shit [0.864]
Fear	They are building a shell command on a server, combining that with user input, and then executing that in a shell on the client. #shudder [0.482]	Let's hope the ct scan gives us some answers on this lump today #nervous [0.818]
Joy	'You're here to feed me. I won't die of starvation,' he said, slightly smiling. I frowned. Panira. Kainis. [0.177]	Omgsh Alexis is sooooo freaking funny on #BachelorInParadise That pizza segment! Plus I love her and Jazzy's friendship! [0.813]
Sadness	I will not fall to the dark side [0.208]	That moment when you look back and realise you've been a #selfish #horrible #judgemental person. #FeelingAshamed [0.909]

Table 1: Examples of tweets with their emotion intensity scores from the task dataset.

notated tweets across four emotions of anger, fear, joy and sadness, which formed the training dataset for the present SemEval-2018 Task 1 dataset (Mohammad et al., 2018). Table 1 provides a few examples from the dataset.

The ability to detect the degree or intensity of emotions is beneficial to many AI applications. For example, a virtual service assistant would be able to employ more appropriate response strategy when a high-intensity anger or frustration is sensed from its customer, as compared to respond monotonically in normal dialogues. Customer relationship management systems can be more targeted by engaging customers who express high degrees of joy or excitement with their products and services. Homecare robots, empowered with the ability to recognize high-intensity grief or distress, would be less likely to miss the opportunity to alert professional human care givers.

In this paper, we discuss our approach to address this emotion intensity detection task, with a focus on the use of and experiments with affective lexicons. In the following sections, we introduce our in-house developed Emotion Intensity lexicons, and compare the performance of feature sets derived from various affective lexicons as well as parts-of-speech, n-grams and word embedding with SVM-based classifier.

2 Emotion Intensity and Affective Lexicons

In its simplest form, emotion intensity refers to the degree or amount of an emotion (Mohammad and Bravo-Marquez, 2017). A basic feature of emotion intensity would be the use of quantifier words. For example, one may indicate that he or she is a *bit* annoyed, *very* pissed off, or *extremely* angry. On the other hand, one may also say that he or she feels *angry*, *livid*, or *furious*. Without quantifier

words, emotion words in itself are salient features indicating the intensity of emotions.

In 2016, we started in-house efforts to develop a multidimensional affective lexicon that computationally captures and distinguishes different psychological and linguistic meanings associated with each emotion-related word. Our initial version of Emotion Intensity (EI) Lexicon is a collection of 3,204 emotion-related English words, common emoticons and Internet slangs labelled in strength and intensity dimensions (as used in Gupta and Yang 2017). In the beginning, the rationale underlying our work centered on the fact that human emotions can be characterized using two fundamental dimensions: the dimension of evaluation *strength* in that an expression would have different levels of pleasantness or unpleasantness (Osgood et al., 1957), and the dimension of *intensity* (Shaver et al., 1987) which concerns about—and what Osgood et al. (1957) originally called as—motivational “potency” and physical “activity”¹. By developing a lexicon that distinguishes strength and intensity, anger-based expressions (high in potency), for example, can be differentiated from equally unpleasant, sadness-based expressions (low in potency).

In EI, each lexicon item is coded with strength and intensity scores in the range of [-3, -2, -1, 0, 1, 2, 3]. For example, items such as “excited”, “astonished” and “thrill” are coded as “3” (high-intensity, positive); items such as “thank”, “cooperative”, “concern”, “:)” and “:d” are coded as “1” (low-intensity, positive); items such as “sorry”, “agh” and “:/” are coded as “-2” (medium-intensity, negative); items such as “hate”, “resented”, “D:” are coded as “-3” (high-intensity, negative); words such as “great”, “haze”,

¹For illustration, the feeling of depression is very negative in evaluation strength, but it may not be highly intense.

“fulfill”, “sick” and “sleepy” are coded as “0” as they are related to emotions by reflecting external, bodily, physical or cognitive conditions, but are not “genuine” emotions (Ortony et al., 1987; Clore et al., 1987).

In Gupta and Yang (2017), we explored the use of the Emotion Intensity (EI) Lexicon and found it helpful in enhancing sarcasm detection and sentiment analysis. Encouraged by its effectiveness, we continue to develop and use the lexicon by adding more psychologically meaningful affective dimensions. We consider three more dimensions: the “basic” emotion categories (Shaver et al., 1987; Ekman, 1973) including fear, anger, sadness, joy, love and surprise, *fine-grained emotion categories* (as summarized in Robinson 2009) including finer emotions such as joy-contentment, joy-cheerfulness, joy-excitement, and *psychological conditions* including affective condition, cognitive condition, physical & bodily state and external condition (Clore et al., 1987). In addition, we also add a *levels of polarity* dimension to reflect if a word is more uni-polarized (e.g., “angry” and “careless” are *definitely* negative) or more bi-polarized (e.g., “surprised” and “sympathetic” may imply both positive and negative feelings). These new considerations contribute to forming the Enhanced Emotion Intensity (E2I) Lexicon. The following table (Table 2) presents the properties of our lexicon in the context of five affective lexicons which were shown to be useful in prior sentiment and emotion analysis research.

3 CrystalFeel System

Focusing on features design and experiments, we employed SVM as the main classifier for the CrystalFeel system. In terms of features, we considered two broad categories: affective lexicon based features, and non-affective lexicon based features.

²<https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#lexicon>

³http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=6010

⁴<http://sentistrength.wlv.ac.uk>

⁵<http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm>

⁶<http://saifmohammad.com/WebPages/AccessResource.htm>

⁷<http://www.crystalace.socialanalyticsplus.net>

⁸The construction of the full Enhanced Emotion Intensity Lexicon is under development of a research manuscript. The lexicon will be released to the research community.

3.1 Affective Lexicons-Based Feature Sets

Following the discussion in Section 2, seven sets of affective lexicons based features were extracted for the experiments:

- OL (6 features): counts of positive (+ive) & negative (−ive) words (2 features), order of +ive & −ive words (1 feature), count of both +ive & −ive words (1 feature), start positions of first occurrence of +ive & −ive words (2 features)
- SS (16 features): counts of +ive & −ive words (2), order of +ive & −ive words (1), count of both +ive & −ive words (1), start positions of first occurrence of +ive & −ive words (2), count of words of different strengths of -5 to 4 (10)
- AFFIN (17 features): counts of +ive & −ive words (2), order of +ive & −ive words (1), count of both +ive & −ive words (1), start positions of first occurrence of +ive & −ive words (2), count of words of different strengths of -5 to 5 (11)
- NRC-EmoLex (14 features): counts of +ive & −ive words (2), order of +ive & −ive words (1), count of both +ive & −ive words (1), start positions of first occurrence of +ive & −ive words (2), count of words belonging to 8 emotions (8)
- NRC-Hash-Emo (14 features): counts of +ive & −ive words (2), order of +ive & −ive words (1), count of both +ive & −ive words (1), start positions of first occurrence of +ive & −ive words (2), total intensity of words belonging to 8 emotions (8)
- EI (24 features): counts of +ive & −ive words (2), order of +ive & −ive words (1), count of both +ive & −ive words (1), start positions of first occurrence of +ive & negative words (2), counts of words holding three strengths of 1 to 3 (3), count of words holding three intensities of 1 to 3 (3), counts of positive & negative words holding three strengths of 1 to 3 (6), counts of positive & negative words holding three intensities of 1 to 3 (6)
- E2I (115 features, including 24 EI features): counts of +ive & −ive words (2), order of +ive & −ive words (1), count of both +ive

		Lexicon Size	Lexical/Words Coverage	Affective Dimension Size	Affective Dimension Coverage	Construction Method
1.	Opinion Lexicon (OL) ² (Hu and Liu, 2004)	6,789	standard English, informal English	1	sentiment/valence	Automatic
2.	AFINN ³ (Nielsen, 2011)	2,477	standard English, informal English	2	sentiment/valence, strength	Manual (Expert Annotation)
3.	SentiStrength (SS) ⁴ (Thelwall et al., 2012)	2,503	standard English, informal English, emoticons, idioms	2	sentiment/valence, strength	Manual (Expert Annotation)
4.	NRC Word-Emotion Association Lexicon (NRC-Emo-Lex) ⁵ (Mohammad and Turney, 2010, 2013)	14,182	standard English	2	sentiment/valence, emotion category (8 emotions)	Manual (Crowd-sourced Annotation)
5.	NRC Hashtag Emotion Association Lexicon (NRC-Hash-Emo) ⁶ (Mohammad, 2012; Mohammad and Kiritchenko, 2015)	16,862	informal English	2	emotion category (8 emotions), association (real-valued numeric score)	Automatic
6.	Emotion Intensity (EI) Lexicon ⁷ (Gupta and Yang, 2017)	3,204	standard English, informal English, emoticons	3	sentiment/valence, strength, intensity	Manual (Expert Annotation)
7.	Enhanced Emotion Intensity (E2I) Lexicon ⁸	3,446	standard English, informal English, emoticons, idioms	7	sentiment/valence, strength, intensity, emotion category (6 basic emotions), fine-grained emotion category (31 fine emotions), psychological condition (4 conditions), polarity condition (4 conditions)	Manual (Expert Annotation)

Table 2: Affective lexicons used in our emotion intensity analysis experiments.

& -ive words (1), start positions of first occurrence of +ive & -ive words (2), counts of words holding three strengths of 1 to 3 (3), count of words holding three intensities of 1 to 3 (3), counts of words belonging to 6 emotions and unspecific words (7), counts of words belonging to 31 emotions and unspecific words (32), counts of words belonging to 4 psychological conditions (4), counts of words belonging to 4 polarity conditions (4), pairwise intersection features across all the dimensions (56)

3.2 N-grams, POS, Word Counts, and Word Embedding Feature Sets

In addition to affective lexicons, we extracted 25-dimensional Tweet part-of-speech (POS) features (Owoputi et al., 2013) for each tweet. Furthermore, as emotion intensity is likely to be associated with the total tweet length and use of capital letters, we added the word counts (WC) features set including counting of total words and counting of uppercase letters. We extracted n-

grams in the same way as in our earlier work (Gupta and Yang, 2017). Lastly, we used FastText (Joulin et al., 2016) to convert the tweets into a 100-dimensional feature vectors. To train FastText model, we downloaded close to 8 million tweets using Twitter Streaming API. In summary:

- POS (25 features): Counts of proper nouns, verbs, conjunctions, adjectives, hashtags, emoticons, urls, punctuations etc. computed using TweetPOS tagger
- WC (2 features): Counts of uppercase letters, text length
- N-grams (12,650 features): uni-grams & bi-grams
- Word Embedding (100 features): word embedding features computed using FastText

3.3 Feature Experiments

Based on the setup, we use the official SemEval-2018 Task 1 training and development datasets to

	Pearson correlations (r)				
	Anger	Fear	Joy	Sadness	Avg.
Individual Affective Lexicons Feature Sets					
OL	0.364	0.426	0.538	0.383	0.428
SS	0.290	0.484	0.445	0.402	0.405
AFFIN	0.344	0.383	0.508	0.440	0.419
NRC-EmoLex	0.216	0.414	0.300	0.301	0.308
NRC-Hash-Emo	0.453	0.492	0.477	0.359	0.445
EI	0.195	0.378	0.368	0.375	0.329
E2I	0.362	0.475	0.458	0.531	0.456
Combined Affective Lexicon Features Sets					
OL + SS	0.300	0.508	0.545	0.457	0.453
OL + SS + AFFIN	0.358	0.497	0.587	0.525	0.492
OL + SS + AFFIN + NRC-EL	0.353	0.544	0.576	0.540	0.503
OL + SS + AFFIN + NRC-EL + NRC-H-E	0.528	0.629	0.627	0.585	0.592
OL + SS + AFFIN + NRC-EL + NRC-H-E + E2I	0.544	0.629	0.626	0.622	0.605
Other Individual Feature Sets					
POS	0.136	0.123	0.070	0.223	0.138
WC	-0.005	0.106	0.069	0.070	0.060
N-grams	0.378	0.608	0.497	0.504	0.497
Word Embedding	0.611	0.557	0.585	0.580	0.583
All Features Sets					
All Features	0.702	0.689	0.666	0.689	0.686

Table 3: Results of feature experiments trained and tested on SemEval-2018 training and development data (highest results in each features sets group are bolded).

train and test the performance of various individual and combined feature conditions. The results are presented in Table 3.

Among individual lexicon based feature sets, features derived from E2I alone led to highest macro-averaged Pearson correlations ($r = 0.456$). (Note that r ranges from -1 to 1 where -1 means perfect reverse correlation and 1 means perfect correlation; a random algorithm gives close to 0 .) The performances of NRC-Hash-Emo and OL came closely as second ($r = 0.445$) and third ($r = 0.428$). Interestingly, on specific emotions, E2I’s advantage ($r = 0.531$) on the prediction of sadness is significantly greater than the second highest prediction of sadness from AFFIN derived features ($r = 0.440$). NRC-Hash-Emo led to highest results for anger ($r = 0.453$) and fear ($r = 0.492$), and OL features led to the highest value for joy ($r = 0.538$).

For the combined affective lexicon features settings, we observed that there was a tendency for each feature set to result in additional advantage (e.g., combining OL + SS features with OL or SS features alone) on the macro-averaged scores, suggesting the complementarity across these lexicons. Combining all the lexicons resulted in a large improvement (avg. $r = 0.605$).

Among non-affective lexicons based features sets, word embedding features obtained the best result ($r = 0.583$). Except for predicting fear,

in which n-grams performed better ($r = 0.608$), word embedding’s advantage also held for predicting anger ($r = 0.611$), joy ($r = 0.585$) and sadness ($r = 0.580$).

Finally, we combined all the lexicon-based features (with small variations⁹ from the individual experiment conditions) and non-lexicon based features. This “all-features” condition resulted in the highest performance for avg. Pearson correlation ($r = 0.684$) and individual correlations for all four emotions. The all-features setting was used for the CrystalFeel system for gold test data.

3.4 Word-level and Message-level Analysis

So to what extent are emotion words from affective lexicons indicative of emotion intensity in tweets at the message level? To explore this question, we ran correlation analysis by calculating bivariate Pearson correlation coefficients between each feature derived from the affective lexicons and the emotion intensity ground truth labels. Figure 1 shows the results.

The analysis indicated several interesting patterns related to the usefulness of lexicon dimensions. First, the sentiment/valence dimension of affective lexicons were generally useful, as the counts of +ive and -ive words (regardless the source of the lexicons) showed up in top ten fea-

⁹The variations are not including two features of start positions of first occurrence of +ive & -ive words (2) for NRC-EmoLex and NRC-Hash-Emo.

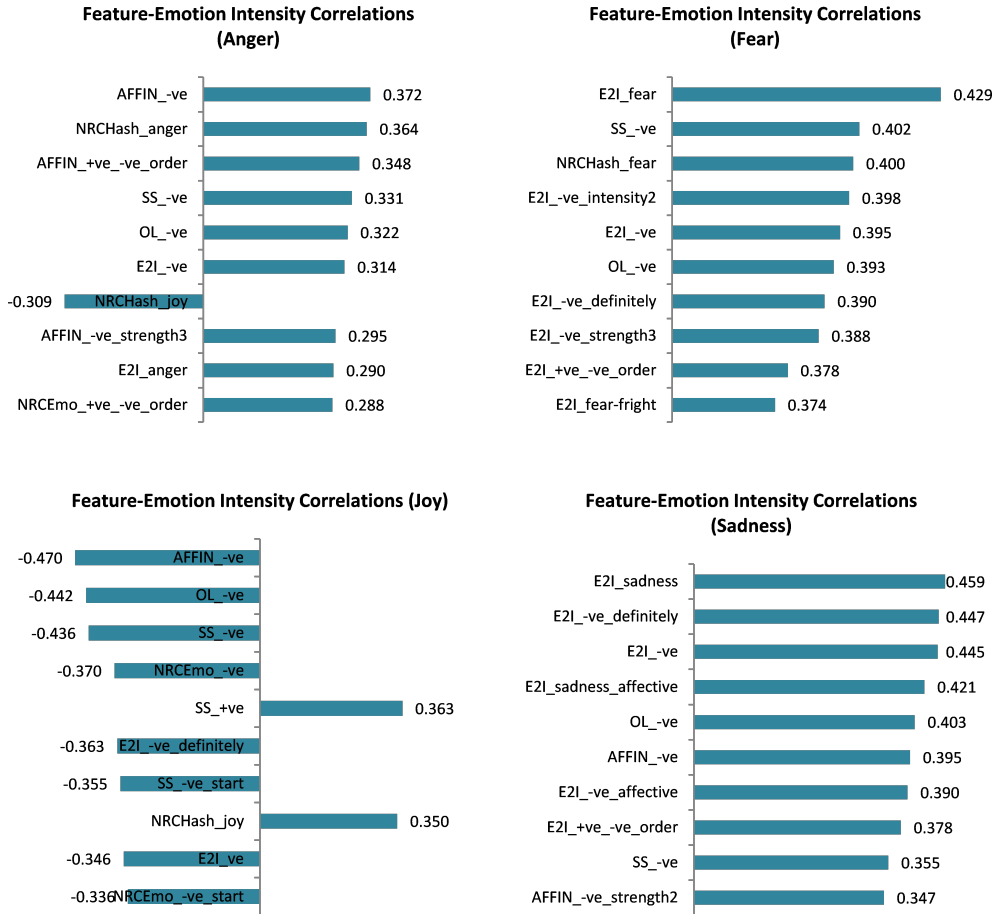


Figure 1: Top ten features with highest bivariate feature-emotion intensity correlations across the four emotion datasets.

	Pearson correlations (r)				
	Anger	Fear	Joy	Sadness	Avg.
Random_Baseline	-0.008	-0.018	-0.058	0.024	-0.008
SVM_Unigrams_Baseline	0.520	0.526	0.575	0.525	0.520
CrystalFeel	0.740	0.700	0.708	0.720	0.717

Table 4: Evaluation Results on subtask 1 emotion intensity regression (EI-reg).

	Pearson correlations (r)				
	Anger	Fear	Joy	Sadness	Avg.
Random_Baseline	-0.008	-0.018	0.024	-0.058	-0.008
SVM_Unigrams_Baseline	0.520	0.526	0.525	0.575	0.520
CrystalFeel	0.576	0.466	0.540	0.538	0.530

Table 5: Evaluation Results on subtask 2 emotion intensity ordinal classification (EI-oc).

tures for all four emotions. Second, features derived from the advanced dimensions of the affective lexicons appeared across the emotions too, suggesting the meaningfulness of lexicons distinguishing finer dimensions of psycholinguistic properties of words. Specifically, for example, on strength dimension, count of negative words with strength 3 from AFFIN lexicon (AFFIN_-ive_strength3) is positively correlated with anger intensity in tweets ($r = 0.295$). On emotion

category dimension, counts of anger words from NRC-Hash-Emo (NRCHash_anger) and from E2I (E2I_anger) are positively associated with anger intensity in tweets. Third, more fine-grained dimensions from E2I are shown as top individual features correlated with anger and sadness intensities and in particular with sadness intensity. For example, count of fear-fright words from E2I (E2I_fear-fright) is highly correlated with fear intensity ($r = 0.374$) and count of sadness words

as genuine emotions (E2I_sadness_affective) is highly correlated with sadness intensity ($r = 0.421$).

Furthermore, the results revealed interesting word-level and message-level feature associations across the four emotions. While the top features for intensities of anger, fear and sadness in tweets (9 or 10 out of 10 top features) are *positive* associations with the presence and higher amount of negative or emotion-specific words, the top features for intensity of joy (7 out of 10 top features) are *negative* associations with the absence and lower amount of negative words. It deserves future research to further investigate these patterns and to cross examine these patterns in other datasets.

	Pearson correlations (r)
	Valence
Random_Baseline	0.031
SVM_Unigrams_Baseline	0.585
CrystalFeel	0.816

Table 6: Evaluation Results on subtask 3 valence/sentiment intensity regression (V-reg).

	Pearson	Kappa
	Valence	Valence
Random_Baseline	-0.010	-0.010
SVM_Unigrams_Baseline	0.509	0.504
CrystalFeel	0.652	0.637

Table 7: Evaluation Results on subtask 4 valence/sentiment ordinal classification (V-oc).

	Accuracy	Micro-avg F_1	Macro-avg F_1
Random_Baseline	0.185	0.307	0.285
SVM_Unigrams_Baseline	0.442	0.570	0.443
CrystalFeel	0.468	0.601	0.522

Table 8: Evaluation Results on subtask 5 multi-label emotion classification (E-c).

4 Results

We evaluated the CrystalFeel system using gold test datasets provided by SemEval-2018 Task 1 (Mohammad et al., 2018). Besides testing the main task of emotion intensity, since it is our primary interest, we have also participated in all other subtasks. In all subtasks, CrystalFeel system outperformed the baseline set by the task organizer. Tables 4-8 summarize the final results.

5 Conclusion

This paper describes CrystalFeel system which is capable of predicting the intensity of emotions associated with a Twitter message. The results of the feature experiments supported the usefulness of our in-house developed EI & E2I lexicons as a new manually constructed lexicon on a relatively small number of lexicon items. In addition, the lexicon also aided us to understand the different patterns of associations between emotion-specific words and emotion-specific intensities at the tweets/messages level. Based on the current analysis, it appeared that our approach possesses a special advantage in understanding and predicting sadness-specific intensity present in tweets. For the use of classifiers, we focused on using SVM as our machine learning classifier in the present study. We plan to investigate the use of deep learning methods in future work.

Acknowledgment

This research is supported by SERC Strategic Fund from Science and Engineering Research Council (SERC), A*STAR (project no. a1718g0046). The authors thank Andrew Ortony for his valuable comments on emotions and emotion intensity and the Digital Emotions team for helpful discussions. We are grateful for the help from Nur Atiqah Othman for her proofreading and enhancement on the clarity of the paper.

References

- Cecilia Ovesdotter Alm, Dan Roth, and Richard Sproat. 2005. Emotions from text: machine learning for text-based emotion prediction. *International conference on Human Language Technology and Empirical Methods in Natural Language Processing* pages 579–586.
- Saima Aman and Stan Szpakowicz. 2007. Identifying expressions of emotion in text. *International Conference on Text, Speech and Dialogue* pages 196–205.
- Gerald L. Clore, Andrew Ortony, and Mark A. Foss. 1987. The psychological foundations of the affective lexicon. *Journal of Personality and Social Psychology* 53(4):751–766.
- Paul Ekman. 1973. cross-cultural studies of facial expression. *Darwin and facial expression: A century of research in review* .

- Nico H. Frijda, Andrew Ortony, Joep Sonnemans, and Gerald Clore. 1992. The complexity of intensity: Issues concerning the structure of emotion intensity. *Personality and Social Psychology Review* 13:60–89.
- Raj Kumar Gupta and Yinping Yang. 2017. CrystalNest at SemEval-2017 Task 4: Using sarcasm detection for enhancing sentiment classification and quantification. *International Workshop on Semantic Evaluation (SemEval 2017)* .
- Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. *ACM SIGKDD* pages 168–177.
- Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2016. Bag of tricks for efficient text classification. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics* 2:427–431.
- Saif M. Mohammad. 2012. #emotional tweets. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics (*Sem)* .
- Saif M. Mohammad and Felipe Bravo-Marquez. 2017. Emotion intensities in tweets. In *Proceedings of the Sixth Joint Conference on Lexical and Computational Semantics* .
- Saif M. Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. 2018. Semeval-2018 task 1: Affect in tweets. *SemEval* .
- Saif M. Mohammad and Svetlana Kiritchenko. 2015. Using hashtags to capture fine emotion categories from tweets. *Computational Intelligence* 31(2):301–326.
- Saif M. Mohammad and Peter Turney. 2010. Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In *Proceedings of the NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text* .
- Saif M. Mohammad and Peter Turney. 2013. Crowdsourcing a word-emotion association lexicon. *Computational Intelligence* 29(3):436–465.
- Finn Arup Nielsen. 2011. AFINN. *Informatics and Mathematical Modelling, Technical University of Denmark* .
- Andrew Ortony, Gerald L. Clore, and Mark A. Foss. 1987. The referential structure of the affective lexicon. *Cognitive Science* 11(3):341–364.
- Charles Egerton Osgood, George J. Suci, and Percy H. Tannenbaum. 1957. The measurement of meaning. *University of Illinois Press* .
- Olutobi Owoputi, Brendan O’Connor, Chris Dyer, Kevin Gimpel, Nathan Schneider, and Noah A. Smith. 2013. Improved part-of-speech tagging for online conversational text with word clusters. *NAACL: HLT* pages 380–390.
- Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *New Publishers Inc* .
- David L. Robinson. 2009. Brain function, emotional experience and personality. *The Netherlands Journal of Psychology* pages 152–167.
- Phillip Shaver, Judith Schwartz, Donald Kirson, and Cary O’Connor. 1987. Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology* 52(6):1061–1086.
- Richard Socher, Alex Perelygin, Jean Y. Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng, and Christopher Potts. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. *EMNLP* .
- Mike Thelwall, Kevan Buckley, and Georgios Paltoglou. 2012. Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology* 63(1):163–173.
- Shiyang Wen and Xiaojun Wan. 2014. Emotion classification in microblog texts using class sequential rules. *AAAI Conference on Artificial Intelligence* pages 187–193.