

Exploring Fine-Grained Emotion Detection in Tweets

Jasy Liew Suet Yan, Howard R. Turtle

School of Information Studies, Syracuse University
Syracuse, New York, USA
jliewsue@syr.edu, turtle@syr.edu

Abstract

We examine if common machine learning techniques known to perform well in coarse-grained emotion and sentiment classification can also be applied successfully on a set of fine-grained emotion categories. We first describe the grounded theory approach used to develop a corpus of 5,553 tweets manually annotated with 28 emotion categories. From our preliminary experiments, we have identified two machine learning algorithms that perform well in this emotion classification task and demonstrated that it is feasible to train classifiers to detect 28 emotion categories without a huge drop in performance compared to coarser-grained classification schemes.

1 Introduction

In sentiment analysis, emotion provides a promising direction for fine-grained analysis of subjective content (Aman & Szpakowicz, 2008; Chaumartin, 2007). Sentiment analysis is mainly focused on detecting the subjectivity (objective or subjective) (Wiebe et al., 2004) or semantic orientation (positive or negative) (Agarwal et al., 2011; Kouloumpis et al., 2011; Pak & Paroubek, 2010; Pang et al., 2002) of a unit of text (i.e., coarse-grained classification schemes) rather than a specific emotion. Often times, knowing exactly how one reacts emotionally towards a particular entity, topic or event does matter (Mohammad et al., 2014). For example, while anger and sadness are both negative emotions, distinguishing between them can be important so businesses can filter out angry customers and respond to them effectively.

Automatic emotion detection on Twitter presents a different set of challenges because tweets exhibit a unique set of characteristics that are not shared by other types of text. Unlike traditional text, tweets consist of short messages expressed within the limit of 140 characters. Due to the length limitation, language used to express emotions in tweets differs significantly from that found in longer documents (e.g., blogs, news, and stories). Language use on Twitter is also typically informal (Eisenstein, 2013; Baldwin et al., 2013). It is common for abbreviations, acronyms, emoticons, unusual orthographic elements, slang, and misspellings to occur in these short messages. On top of that, retweets (i.e., propagating messages of other users), referring to @username when responding to another user’s tweet, and using #hashtags to represent topics are prevalent in tweets. Even though users are restricted to post only 140 characters per tweet, it is not uncommon to find a tweet containing more than one emotion.

Emotion cues are not limited to only emotion words such as *happy*, *amused*, *sad*, *miserable*, *scared*, etc. People use a variety of ways to express a wide range of emotions. For instance, a person expressing happiness may use the emotion word “happy” (Example 1), the interjection “woop” (Example 2), the emoticon “:)” (Example 3) or the emoji “😊” (Example 4).

Example 1: “I can now finally say I am at a place in my life where I am happy with who am and the stuff I have coming for me in the future #blessed” [Happiness]

Example 2: “its midnight and i am eating a lion bar woop” [Happiness]

Example 3: “Enjoying a night of #Dexter with @DomoniqueP07 :)” [**Happiness**]

Example 4: “The wait is almost over LA, will be out in just a little! 😊😊😊😊😊” [**Happiness**]

In addition to explicit expressions of emotion, users on Twitter also express their emotions in figurative forms through the use of idiomatic expressions (Example 5), similes (Example 6), metaphors (Example 7) or other descriptors (Example 8). In these figurative expressions of emotion, each word if treated individually does not directly convey any emotion. When combined together and, depending on the context of use, they act as implicit indicators of emotion. Automatic emotion detectors that rely solely on the recognition of emotion words will likely fail to recognize the emotions conveyed in these examples.

Example 5: “@ter2459 it was!!! I am still on cloud nine! I say and watched them for over two hours. I couldn't leave! They are incredible!” [**Happiness**]

Example 6: “Getting one of these bad boys in your cereal box and feeling like your day simply couldn't get any better <http://t.co/Fae9EjyN61>” [**Happiness**]

Example 7: “Loving the #IKEAHomeTour décor #ideas! Between the showroom and the catalog I am in heaven” [**Happiness**]

Example 8: “I did an adult thing by buying stylish bed sheets and not fucking it up when setting them up. *cracks beer open*” [**Happiness**]

The occurrence of an emotion word in a tweet does not always indicate the tweeter's emotion. The emotion word “happy” in Example 9 is not used to describe how the tweeter feels about the tune but is instead used to characterize the affective quality or affective property of the tune (Russell, 2003; Zhang, 2013). The tweeter attributes a happy quality to the tune but is in fact expressing anger towards the “happy” tune. Similarly, #Happiness in Example 10 is part of a book's title so the emotion word hashtag functions as a topic more than an expression or description of an individual's emotion. The common practice of using emotion word hashtags to retrieve self-annotated examples as ground truth to build emotion classifiers, a method known as “distant supervision” (Hasan et al., 2014; Mohammad, 2012; Mohammad & Ki-

ritchenko, 2014; Wang et al., 2012), is susceptible to this weakness.

Example 9: “@Anjijade I was at this party on the weekend, that happy tune was played endlessly, really not my stuff, it was like the cure's torture ha” [**Anger**]

Example 10: “Hear Carrie Goodwiler's audition for the audio version of my book #Happiness & Honey on #SoundCloud” [**No Emotion**]

These challenges associated with detecting fine-grained emotion expressions in tweets remain a virgin territory that has not been thoroughly explored. To start addressing some off these challenges, we present a manually-annotated tweet corpus that captures a diversity of emotion expressions at a fine-grained level. We describe the grounded theory approach used to develop a corpus of 5,553 tweets manually annotated with 28 emotion categories. The corpus captures a variety of explicit and implicit emotion expressions for these 28 emotion categories, including the examples described above.

Using this carefully curated gold standard corpus, we report our preliminary efforts to train and evaluate machine learning models for emotion classification. We examine if common machine learning techniques known to perform well in coarse-grained emotion and sentiment classification can also be applied successfully on this set of fine-grained emotion categories. The contributions of this paper are two-fold:

- a) Identifying machine learning algorithms that generally perform well at classifying the 28 emotion categories in the corpus and comparing them to baselines
- b) Comparing the machine learning performance of fine-grained to coarse-grained emotion classification

2 Empirical Study

2.1 Corpus

The corpus contains 5,553 tweets and is developed using small-scale content analysis. To ensure that the tweets included in the corpus are representative of the population on Twitter, we employed four sampling strategies: randomly sampling tweets retrieved using common stopwords (RANDOM: 1450 tweets), sampling using topical hashtags

(TOPIC: 1310 tweets), sampling using @usernames of US Senators (SEN-USER: 1493 tweets) and sampling using @usernames of average users randomly selected from Twitter (AVG-USER: 1300 tweets). Tweets were sampled from the Twitter API and two publicly available datasets: 1) the SemEval 2014 tweet data set (Nakov et al., 2013; Rosenthal et al., 2014), and 2) the 2012 US presidential elections data set (Mohammad et al., 2014). The proportion of tweets from each of the four samples is roughly balanced.

The corpus was annotated by graduate students who were interested in undertaking the task as part of a class project (e.g., Natural Language Processing course) or to gain research experience in content analysis (e.g., independent study). A total of 18 annotators worked on the annotation task over a period of ten months. Annotators were first instructed to annotate the valence of a tweet. Emotion valence can be positive, negative or neutral. Positive emotions are evoked by events causing one to express pleasure (e.g., happy, relaxed, fascination, love) while negative emotions are evoked by events causing one to express displeasure (e.g., anger, fear, sad). Emotions that were neither positive nor negative were considered to be neutral (e.g. surprise). Valence was useful to help annotators distinguish between tweets that contained emotion and those that did not.

To uncover a set of emotion categories from the tweets, we used an adapted grounded theory approach developed by Glaser & Strauss (1967) for the purpose of building theory that emerges from the data. Using this approach, annotators were not given a predefined set of labels for emotion category. Instead, the emotion categories were formed inductively based on the emotion tags or labels suggested by annotators. Annotators were required to identify emotion tag when valence for a tweet was labeled as either “Positive”, “Negative” or “Neutral”. For emotion tag, annotators were instructed to assign an emotion label that best described the overall emotion expressed in a tweet. In cases where a tweet contained multiple emotions, annotators were asked to first identify the primary emotion expressed in the tweet, and then also include the other emotions observed.

The annotation task was conducted in an iterative fashion. In the first iteration, also referred to as

the training round, all annotators annotated the same sample of 300 tweets from the SEN-USER sample. Annotators were expected to achieve at least 70% pairwise agreement for valence with the primary researcher in order to move forward. The annotators achieved a mean pairwise agreement of 82% with the researcher. Upon passing the training round, annotators were assigned to annotate at least 1,000 tweets from one of the four samples (RANDOM, TOPIC, AVG-USER or SEN-USER) in subsequent iterations. Every week, annotators worked independently on annotating a subset of 150 – 200 tweets but met with the researcher in groups to discuss disagreements, and 100% agreement for valence and emotion tag was achieved after discussion. In these weekly meetings, the researcher also facilitated the discussions among annotators working on the same sample to merge, remove, and refine suggested emotion tags.

Annotators suggested a total 246 distinct emotion tags. To group the emotion tags into categories, annotators were asked to perform a card sorting exercise in different teams to group emotion tags that are variants of the same root word or semantically similar into the same category. Annotators were divided into 5 teams, and each team received a pack of 1’ x 5’ cards containing only the emotion tags used by the all members in their respective teams. This task organized the emotion tags into 48 emotion categories.

To refine the emotion categories, we collected pleasure and arousal ratings for each emotion category name from Amazon Mechanical Turk (AMT). Based on 76 usable responses, the emotion category names were mapped on a two-dimensional plot. Emotion categories that were closely clustered together on the plot and semantically related to one another were further merged resulting in a final set of 28 emotion categories. Finally, all emotion category labels in the corpus were systematically replaced by the appropriate 28 emotion category labels. Overall, annotators achieved Krippendorff’s $\alpha = 0.61$ for valence and $\alpha = 0.50$ for the set of 28 emotion categories. Each tweet was assigned gold labels for valence and emotion category.

2.2 Emotion Distributions

This section describes the distribution of gold labels for three emotion class structures: 1) emo-

tion/non-emotion, 2) valence, and 3) 28 emotion categories. As shown in Table 1, the overall distribution between tweets containing emotion and those that do not is roughly balanced. Slightly over half of the tweets (53%) contain emotion.

Class	n	%
Emotion	2916	53
Non-Emotion	2637	47
Total	5553	100

Table 1: Distribution of tweets with emotion and non-emotion

Class	n	%
Positive	1840	33
Negative	744	13
Neutral	155	3
Multiple Valence	177	3

Table 2: Distribution of tweets for emotion valence

Class	n	%
Admiration	132	2.38
Amusement	216	3.89
Anger	409	7.37
Boredom	10	0.18
Confidence	15	0.27
Curiosity	27	0.49
Desperation	7	0.13
Doubt	44	0.79
Excitement	228	4.11
Exhaustion	7	0.13
Fascination	47	0.85
Fear	65	1.17
Gratitude	176	3.17
Happiness	697	12.55
Hate	52	0.94
Hope	132	2.38
Indifference	24	0.43
Inspiration	16	0.29
Jealousy	5	0.09
Longing	27	0.49
Love	172	3.10
Pride	70	1.26
Regret	40	0.72
Relaxed	25	0.45
Sadness	137	2.47
Shame	22	0.40
Surprise	83	1.49
Sympathy	31	0.56

Table 3: Distribution of tweets for 28 emotion categories

The class distribution becomes more unbalanced with the finer-grained emotion classes, valence (Table 2) and 28 emotion categories (Table 3). For valence, 33% of the tweets containing emotion are

positive, 13% are negative and only 3% are neutral. Emotion classes become even sparser with the 28 emotion categories. The most frequent category is *happiness* (13%) while the least frequent category is *jealousy* (0.09%).

2.3 Machine Learning Experiments

We ran a series of experiments to identify a set of machine learning algorithms that generally perform well for this task. Four machine learning algorithms were found to perform well in this problem space: support vector machines (SVM) (Alm et al., 2005; Aman & Szpakowicz, 2007; Brooks et al., 2013; Cherry et al., 2012), Bayesian networks (Sohn et al., 2012; Strapparava & Mihalcea, 2008), decision trees (Hasan et al., 2014), and k-nearest neighbor (KNN) (Hasan et al., 2014; Holzman & Pottenger, 2003). The features were held constant across different classifiers in the candidate set. As a starting point, a unigram (i.e., bag-of-words) model, which has been shown to work reasonably well for text classification in sentiment analysis (Pang et al., 2002; Salvetti et al., 2006), was chosen. Although limited, the unigram bag-of-words features captures not only emotion words but all words in a tweet, thus increasing the likelihood of the classifiers to handle the figurative expressions of emotion.

We tokenized the text in the corpus and extracted all unique terms as features. We created a custom tokenizer to better handle elements that are common in tweets. In particular, the tokenizer recognizes emoticons, emojis, URLs and HTML encoding. The tokenizer also handles common abbreviations and contractions. Text was encoded in UTF-8 in order to preserve the emojis. We then evaluated the effect of case normalization (i.e., lowercasing), stemming, and a minimum word frequency threshold ($f=1, 3, 5$ and 10) as a means to reduce the number of features. Classifiers were evaluated using 10-fold cross validation.

To make experiments more manageable, we frame the problem as a *multi-class* classification task. Each tweet was assigned to only one emotion label. For tweets with multiple labels, only the primary label (i.e., first label) was assigned to the tweet, and the other labels were ignored. We carried out two sets of experiments. First, we created one single classifier (*multi-class-single: one versus*

one) to distinguish between 29 classes (i.e., 28 emotion categories and *no emotion*). Second, we ran experiments using Weka’s *MultiClassClassifier*, a meta-classifier that mapped a multi-class dataset into multiple two-class classifiers (*multi-class-binary: one versus all*), one for each emotion and one for *no emotion*, thus resulting in a setup with 29 binary classifiers in total. Unfortunately, the *multi-class-binary* setup was not designed to handle instances with multiple labels but it offered a straightforward implementation of multiple binary classifications for preliminary analysis. About 92% of the corpus contained instances with only a single label so overall classification performance is expected to be close to that of a *multi-label* classifier.

3 Evaluation and Results

3.1 Machine Learning Algorithms

We found that the use of stemming, case normalization and applying a word frequency threshold of 3 produced consistently good results.

Classifier	MCS	MCB
BayesNet	0.533	0.574
SVM-SMO	0.571	0.529
J48	0.567	0.520
KNN ($k = 1$)	0.391	0.391

Table 4: Micro-averaged F1 for *multi-class-single* (MCS) and *multi-class-binary* (MCB)

Based on the micro-averaged F1 shown in Table 4, the two machine learning algorithms that yielded the best performance were Sequential Minimal Optimization (SMO), an algorithm for training SVM (Platt, 1998) and Bayesian Networks (BayesNet) (Bouckaert, 1967). The performance ranking differs slightly between the four machine learning algorithms across the two experimental setups with SVM being the top performing classifier in *multi-class-single* while BayesNet in *multi-class-binary*. A more in-depth analysis of the best performing classifier for each emotion category also shows that BayesNet and SVM yield the best performance for over half of the emotion categories.

3.2 Comparison with Baselines

Three baselines are first established as the basis of comparison for all other classifiers.

- Majority-class baseline: The majority-class baseline simply assigns the majority class to each tweet.
- Random baseline: The random baseline classifier predicts a label randomly with no learning involved.
- OneR: OneR is a simple classifier that uses a single feature with minimum error for classification. The classifier generates a set of rules based on this single feature.

Classifier	A	P	R	F1
<i>multi-class-single</i>				
Majority-class	47.4	0.23	0.47	0.31
Random	0.03	0.03	0.03	0.03
OneR	49.8	0.26	0.50	0.34
BayesNet	60.1	0.54	0.60	0.51
SVM-SMO	58.9	0.57	0.59	0.57
<i>multi-class-binary</i>				
Majority-class	47.4	0.23	0.47	0.31
Random	0.03	0.03	0.50	0.06
OneR	51.7	0.56	0.52	0.46
BayesNet	63.0	0.60	0.63	0.57
SVM-SMO	48.9	0.61	0.49	0.53

Table 5: Comparison between best performing models and baselines (A: Accuracy, P: Precision, R: Recall)

We compare the SVM and BayesNet classifiers to the three baselines as shown in Table 5. In terms of accuracy, SVM and BayesNet outperform the majority-class and random baselines in both *multi-class-single* and *multi-class-binary*. BayesNet correctly predicts roughly 60% of the instances while SVM correctly predicts roughly 50%. In terms of F1, SVM and BayesNet exceed the performance of all the three baselines.

3.3 Levels of Granularity

Table 6 shows the performance of classifiers for fine-grained versus coarser-grained class structures across three levels of granularity: 1) emotion presence/absence (2 classes), 2) emotion valence (5 classes) and, 3) emotion category (28 classes). SVM and BayesNet perform significantly better than the majority-class baseline across all three levels of granularity using a flat classification approach. The majority class for valence and emotion category is *none*.

Level	SVM-SMO				BayesNet				Majority-class			
	A	P	R	F1	A	P	R	F1	A	P	R	F1
Emotion Presence: Emotion, None												
<i>multi-class-single</i>	72.7	0.73	0.73	0.73	72.2	0.73	0.72	0.72	52.6	0.28	0.53	0.36
Emotion Valence: Positive, Negative, Neutral, Multiple Valence, None												
<i>multi-class-single</i>	65.5	0.63	0.64	0.63	67.0	0.65	0.67	0.65	47.4	0.23	0.47	0.31
Emotion Category (28-Emo-Cat, None)												
<i>multi-class-single</i>	58.9	0.57	0.59	0.57	60.1	0.54	0.60	0.51	47.4	0.23	0.47	0.31
<i>multi-class-binary</i>	48.9	0.61	0.49	0.53	63.0	0.60	0.63	0.57	47.4	0.23	0.47	0.31

Table 6: Accuracy (A), precision (P), recall (R) and F1 across classification schemes with different levels of granularity

Comparing across the three levels of granularity, better performance is observed when there are fewer classes. For example, a classifier trained to distinguish between 2 classes (*emotion* and *none*) yields higher performance than a classifier trained to distinguish between 29 classes (28 emotion categories and *none*). The drop in classifier performance from coarser to finer levels of granularity is gradual. Note that the performance of a classifier trained to classify 29 classes is not a great deal worse than a classifier dealing with fewer classes (2 or 5). A closer analysis of the F1 per emotion category shows that the classifiers are able to correctly predict some categories better than the others. For instance, SVM and BayesNet achieve F1 greater than 0.7 for *gratitude*. The performance measures in Table 6 are micro averages across all classes. The performance results reported here are intended to show a realistic assessment of machine learning performance in classifying the 28 emotion categories that emerged from the open coding task. We included even the poor performing categories in the computation of the micro averages.

4 Discussion and Conclusion

Automatic fine-grained emotion detection is a challenging task but we have demonstrated that it is feasible to train a classifier to perform decently well in classifying as many as 28 emotion categories. Our 28 emotion categories is an extension to the six to eight emotion categories commonly-used in the state-of-the-art (Alm et al., 2005; Aman & Szpakowicz, 2007; Mohammad, 2012). Some of the 28 emotion categories overlap with those found in existing emotion theories such as Plutchik’s (1962) 24 categories on the wheel of emotion and Shaver et al.’s (2001) tree-structured list of emotions. Existing emotion theories in psychology are not developed specifically based on emotions ex-

pressed in text. Therefore, our emotion categories offer a more fitting framework for the study of emotion in text.

Existing classifiers achieve only moderate performance in detecting emotions in tweets even those trained with a significant amount of data collected using distant supervision (Mohammad, 2012; Roberts et al., 2012; Wang et al., 2012). Our preliminary classifiers trained with less data show results that are comparable to existing coarse-grained classifiers. Results from our preliminary machine learning experiments conclude that SVM and BayesNet classifiers produce consistently good performance for fine-grained emotion classification. Therefore, we plan to continue our machine learning experiment with more sophisticated feature selection strategies, ensemble methods and more balanced training data using both SVM and BayesNet.

There is no stark difference in classifier performance between fine-grained and coarse-grained emotion classes. Classifiers perform poorly for a handful of emotion categories with very low frequency. We will need to generate more positive examples for these classes to improve classifier performance. We plan to add another 10,000 annotated tweets in the corpus to increase the size of training and evaluation data. We will make the emotion corpus available in the future.

We acknowledge that the *multi-class* setup may not be the most suitable implementation of this classification task given that the corpus contains tweets annotated with multiple emotion categories. We chose the *multi-class* setup to simplify the classification task and make the machine learning experiments more manageable in this preliminary stage. We plan to evaluate the effectiveness of these algorithms with *multi-label* classifiers in our future work.

Acknowledgments

We thank the annotators who volunteered in performing the annotation task. We are grateful to Dr. Elizabeth D. Liddy for her insights in the study.

References

- Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment analysis of Twitter data. In *Proceedings of the Workshop on Languages in Social Media* (pp. 30–38). Stroudsburg, PA, USA.
- Alm, C. O., Roth, D., & Sproat, R. (2005). Emotions from text: Machine learning for text-based emotion prediction. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing* (pp. 579–586). Stroudsburg, PA, USA.
- Aman, S., & Szpakowicz, S. (2007). Identifying expressions of emotion in text. In *Text, Speech and Dialogue* (pp. 196–205).
- Aman, S., & Szpakowicz, S. (2008). Using Roget’s thesaurus for fine-grained emotion recognition. In *Proceedings of the Third International Joint Conference on Natural Language Processing* (pp. 296–302).
- Baldwin, T., Cook, P., Lui, M., MacKinlay, A., & Wang, L. (2013). How Noisy Social Media Text, How Diffrent Social Media Sources? In *Proceedings of the 6th International Joint Conference on Natural Language Processing (IJCNLP 2013)* (pp. 356–364). Nagoya, Japan.
- Bouckaert, R. R. (1967). *Bayesian belief networks: From construction to inference*. Universiteit Utrecht, Faculteit Wiskunde en Informatica.
- Brooks, M., Kuksenok, K., Torkildson, M. K., Perry, D., Robinson, J. J., Scott, T. J., Anicello, O., Zukowski, A., Harris, P., & Aragon, C. R. (2013). Statistical affect detection in collaborative chat. Presented at the Conference on Computer Supported Cooperative Work and Social Computing, San Antonio, TX.
- Chaumartin, F. R. (2007). UPAR7: A knowledge-based system for headline sentiment tagging. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)* (pp. 422–425).
- Cherry, C., Mohammad, S. M., & de Bruijn, B. (2012). Binary classifiers and latent sequence models for emotion detection in suicide notes. *Biomedical Informatics Insights*, 5, 147–154.
- Eisenstein, J. (2013). What to do about bad language on the internet. In *Proceedings of NAACL-HLT 2013* (pp. 359–369).
- Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.
- Hasan, M., Agu, E., & Rundensteiner, E. (2014). Using hashtags as labels for supervised learning of emotions in Twitter messages.
- Hasan, M., Rundensteiner, E., & Agu, E. (2014). EMOTEX: Detecting emotions in Twitter messages. Presented at the 2014 ASE BIGDATA/SOCIALCOM/CYBERSECURITY Conference, Stanford University.
- Holzman, L. E., & Pottenger, W. (2003). Classification of emotions in internet chat: An application of machine learning using speech phonemes.
- Kouloumpis, E., Wilson, T., & Moore, J. (2011). Twitter sentiment analysis: The good the bad and the omg! In *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media (ICWSM)* (pp. 538–541).
- Mohammad, S. M. (2012). #Emotional tweets. In *Proceedings of the 1st Joint Conference on Lexical and Computational Semantics* (pp. 246–255). Montreal, QC.
- Mohammad, S. M., & Kiritchenko, S. (2014). Using hashtags to capture fine emotion categories from tweets. *Computational Intelligence*, 301-326.
- Mohammad, S. M., Zhu, X., & Martin, J. (2014). Semantic role labeling of emotions in tweets. In *Proceedings of the ACL 2014 Workshop on Computational Approaches to Subjectivity, Sentiment, and Social Media (WASSA)* (pp. 32–41). Baltimore, MD.
- Nakov, P., Kozareva, Z., Ritter, A., Rosenthal, S., Stoyanov, V., & Wilson, T. (2013). SemEval-2013 task 2: Sentiment analysis in Twitter. In *Proceedings of the 7th International Workshop on Semantic Evaluation (SemEval 2013)* (pp. 312–320).
- Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. In *Seventh International Conference on Language Resources and Evaluation (LREC)* (pp. 1320–1326).
- Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: Sentiment classification using machine learning techniques. In *Proceedings of the ACL-2002 Conference on Empirical Methods in Natural Language Processing* (pp. 79–86). Stroudsburg, PA, USA.
- Platt, J. C. (1998). Fast Training of Support Vector Machines Using Sequential Minimal Optimization - Microsoft Research. In *Advances in Kernel Methods - Support Vector Learning* (pp. 41–65). MIT Press.
- Plutchik, R. (1962). *The Emotions: Facts, theories, and a new model*. New York: Random House.
- Roberts, K., Roach, M. A., Johnson, J., Guthrie, J., & Harabagiu, S. M. (2012). EmpaTweet: Annotating and detecting emotions on Twitter. In *8th International Conference on Language Resources and Evaluation (LREC)* (pp. 3806–3813).

- Rosenthal, S., Nakov, P., Ritter, A., & Stoyanov, V. (2014). Semeval-2014 task 9: Sentiment analysis in Twitter. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)* (pp. 73–80). Dublin, Ireland.
- Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172.
- Salveti, F., Reichenbach, C., & Lewis, S. (2006). Opinion polarity identification of movie reviews. *Computing Attitude and Affect in Text: Theory and Applications* (pp. 303–316).
- Shaver, P., Schwartz, J., Kirson, D., & O'Connor, C. (2001). Emotion knowledge: Further exploration of a prototype approach. In *Emotions in Social Psychology* (pp. 26–56). Psychology Press.
- Sohn, S., Torii, M., Li, D., Waghlikar, K., Wu, S., & Liu, H. (2012). A hybrid approach to sentiment sentence classification in suicide notes. *Biomedical Informatics Insights*, *5*(Suppl. 1), 43–50.
- Strapparava, C., & Mihalcea, R. (2008). Learning to identify emotions in text. In *Proceedings of the 2008 ACM Symposium on Applied Computing* (pp. 1556–1560). New York, USA.
- Wang, W., Chen, L., Thirunarayan, K., & Sheth, A. P. (2012). Harnessing Twitter “big data” for automatic emotion identification. In *2012 International Conference on Privacy, Security, Risk and Trust, and 2012 International Conference on Social Computing* (pp. 587–592).
- Wiebe, J. M., Wilson, T., Bruce, R., Bell, M., & Martin, M. (2004). Learning subjective language. *Computational Linguistics*, *30*(3), 277–308.
- Zhang, P. (2013). The Affective Response Model: A theoretical framework of affective concepts and their relationships in the ICT context. *Management Information Systems Quarterly*, *37*(1), 247–274.