EmpaTweet: Annotating and Detecting Emotions on Twitter

Kirk Roberts, Michael A. Roach, Joseph Johnson, Josh Guthrie, Sanda M. Harabagiu

Human Language Technology Research Institute
University of Texas at Dallas
Richardson TX 75080
{kirk,sanda}@hlt.utdallas.edu

Abstract

The rise of micro-blogging in recent years has resulted in significant access to emotion-laden text. Unlike emotion expressed in other textual sources (e.g., blogs, quotes in newswire, email, product reviews, or even clinical text), micro-blogs differ by (1) placing a strict limit on length, resulting radically in new forms of emotional expression, and (2) encouraging users to express their daily thoughts in real-time, often resulting in far more emotion statements than might normally occur. In this paper, we introduce a corpus collected from Twitter with annotated micro-blog posts (or "tweets") annotated at the tweet-level with seven emotions: ANGER, DISGUST, FEAR, JOY, LOVE, SADNESS, and SURPRISE. We analyze how emotions are distributed in the data we annotated and compare it to the distributions in other emotion-annotated corpora. We also used the annotated corpus to train a classifier that automatically discovers the emotions in tweets. In addition, we present an analysis of the linguistic style used for expressing emotions our corpus. We hope that these observations will lead to the design of novel emotion detection techniques that account for linguistic style and psycholinguistic theories.

Keywords: emotion detection, sentiment analysis, linguistic style

1. Introduction

Micro-blogging services such as Twitter provide researchers with a wealth of information on how individuals communicate with their social network. Unlike more formal methods of communication, micro-blog posts (hereafter, "tweets") frequently reflect the author's opinions and emotional states. For instance, Table 1 shows several recent tweets reflecting on the latest FIFA World Cup. Furthermore, since tweets are restricted to 140 characters, and since they are often written on mobile devices, they express emotions less formally than other publishing platforms.

In this paper, we describe the creation of a corpus of tweets on a variety of popular Twitter topics with their corresponding manually-annotated emotions. Topics were chosen based on our expectation of which emotions will be present in topical tweets in order to get a good distribution of our chosen emotions. Machine learning methods can then be trained on these annotations in order to automatically extract emotions from tweets. Such a system would be useful in understanding users' feelings towards particular products, services, or topics (e.g., companies could determine the distribution of emotions toward their latest product). It would additionally enable emotion-temporal analysis (e.g., tracking the emotions of individuals over time).

In summary, the contribution of this work is three-fold:

1. A publicly available corpus of tweets annotated with seven emotions: ANGER, DISGUST, FEAR, JOY, LOVE, SADNESS, and SURPRISE.

2. Competitive, easily implementable baselines that act as a benchmark for automated approaches using this data and illustrate the overall difficulty of the task.

3. An analysis of the emotional and stylistic distributions of our corpus, including comparisons to other available domains.

The remainder of this paper is organized as follows. Section 2 provides related work with emotion detection and the use of Twitter as a sentiment corpus, as well as related work on linguistic style in social media. Section 3 describes our process for creating the corpus. Section 4 analyzes the resulting corpus and compares it to other emotion corpora. Section 5 presents a supervised baseline for detecting emotions based on our corpus. Section 6 discusses the linguistic style characterizing expressions of emotions. Finally, Section 7 summarizes the conclusion and motivates future work.

2. Related Work

The rise of social media has attracted significant interest in sentiment analysis techniques such as emotion detection and opinion mining (Pang and Lee, 2008). Emotion detection has additionally been applied to other domains such as novels (Mohammad, 2011), e-mail (Mohammad and Yang, 2011), news headlines (Strapparava and Mihalcea, 2008), and suicide notes (Pestian et al., 2011). In regards to micro-blogs, this work has focused primarily on large-scale understanding of sentiment (Pak and Paroubek, 2010; Bollen et al., 2011a) for purposes such as understanding consumer views towards a product or predicting the stock market (Bollen et al., 2011b). However, these methods depend on large numbers of tweets and assume the lexical heuristics used to extract certain types of emotional content are representative of the whole. For instance, Pak and Paroubek (2010) use emoticons (e.g., ":-)" and ":D" for happy, ":-(" and "=(" for sad) as queries to retrieve large amounts of unlabeled data under the assumption that these tweets are representative of all happy and sad tweets. In contrast, our goal is to individually label a smaller number of tweets with a finer-grained set of emotions. This not only would allow for greater range of emotion detection, but ensures our corpus has a greater lexical variety, as we are not limited to training on tweets extracted with lexical heuristics. As

@kingpuyol: downloading #WC2010 videos and just weeping everywhere
@AustinLong1974: Latest #SoccerNomad post is up: bit.ly/qeP3sm La Furia Roja ended years of frustration & I missed the glorious moment #WC2010
@BarryBru: The greed of many within tourism here in #SouthAfrica around #WC2010 has without a doubt hurt our industry. We must regain competitive edge!
@petegravestv: Can't help but draw similarities between this #RWC2011 and the Football #WC2010 - generally poor matches and a dodgy ball!

Table 1: Example tweets related to the 2010 FIFA World Cup.

Topic	Hashtags
Valentine's Day	#valentine #valentines #valentinesday #cupid
Lindsay Lohan	#lohan #lindsaylohan
September 11th	#nineeleven #sept11 #september11 #nine11 #9eleven
2012 U.S. Election	#obama #romney #ronpaul #gingrich #gop #gopdebate #republicandebate #teaparty
Palestinian Statehood	#palestine #palestinestate #palestinestatehood #palestineun #gopalestine #freepalestine
Egyptian riots	#arabspring #tahir #tahrir #egyptianrevolution #egypt
Super Bowl XLV	#superbowlxlv #superbowl
World Cup 2010	#worldcup2010 #wc2010 #worldcup
Christmas	#christmas #xmas #santa #happyholidays
DC/NY earthquake	#earthquake #dcearthquake #eastcoastearthquake
Emmys	#emmy #emmys #emmyaward #emmyawards
Eminem	#eminem #eminemsong
stock market	#stocks #stockmarket #dow #dowjones #sandp #nasdaq #wallstreet #NYSE
Greek bailout	#bailout #greece #greekbailout #eurocrisis #euro

Table 2: Chosen topics and their corresponding hashtags.

pointed out by Mohammad (2011), emotion detection has been shown reliable on large amounts of data using existing techniques, but is unpredictable on small amounts of text such as short sentences or micro-blogs. Since our goal is to analyze individual tweets for their emotional content, instead of a massive number of tweets for overall themes, we therefore require highly accurate training data for extracting emotions from context-poor sources such as Twitter.

The growth of the social web and the corresponding rise in available emotional text over the past several years has led to the development of the *We Feel Fine* emotional search engine (Kamvar and Harris, 2011) that employs web-based art work to collect the world's emotions, with the purpose of helping people better understand themselves and others. Although it uses a "Feelings Indexer", the recognition of feelings and emotions is solely based on hand-crafted regular expressions that uses an emotional lexicon. Although the interfaces of the search engine enable data visualization, this work is just exploiting a computational framework for an infrastructure of emotion data collection. Kim et al. (2011) uses a computational framework for analyzing several aspects of sentiment and emotions expressed in Twitter conversations, prompted by the question "Do you feel what I feel?". Of special interest is the study of influence among Twitter conversation participants, which enable the change in sentiment and emotion. A probabilistic topic model, based on latent Dirichlet Allocation, enables the identification of sentiments and emotions in an un-annotated corpus of Twitter conversations. This work highlights the discovery of emotion shifts among Twitter conversation participants.

Twitter conversations are also the focus of the work reported in Danescu-Niculescu-Mizil et al. (2011). This work describes linguistic style accommodation in Twitter conversations by making use of the Linguistic Inquiry Word Count (LIWC) method (Pennebaker and King, 1999). Their experiments show that the hypothesis of linguistic style accommodation holds in social media conversations.

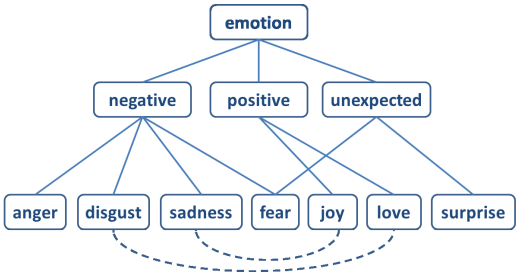

Figure 1: Our emotion ontology for the six Ekman emotions (plus LOVE). Solid lines indicate inheritance, dashed lines indicate opposite.

3. Corpus Creation

We chose seven emotions based on Ekman's six basic emotions and LOVE, which we believed would be commonly found in informal text such as Twitter. We have arranged these seven emotions into an ontology, shown in Figure 1, in order to aid the annotators in understanding how the emotions relate.

We chose 14 topics that we believed would frequently evoke emotion on Twitter. This means that our data is not necessarily representative of Twitter as a whole, but it allows us to guarantee that all seven of our emotions are represented in the data and minimize the number of non-emotion evoking tweets. Since our goal is to enable machine learning-based approaches to the detection of emotions on Twitter, a fairly balanced data set is a reasonable choice. For each topic, we compiled a list of hashtags, shown in Table 2, which are used in tweets to mark top-

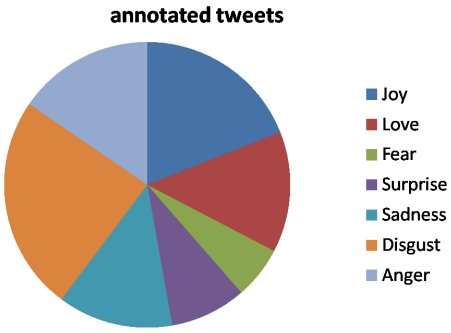

Figure 2: Emotion distribution in annotated Twitter corpus.

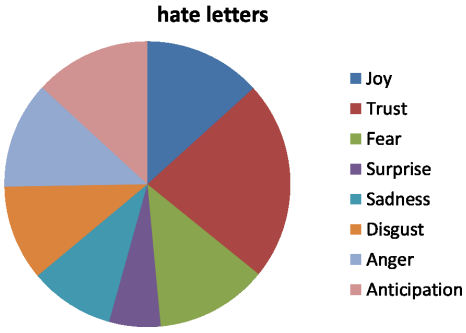

Figure 4: Emotion distribution in hate letters.

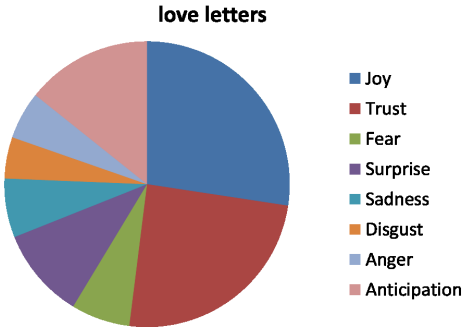

Figure 3: Emotion distribution in love letters.

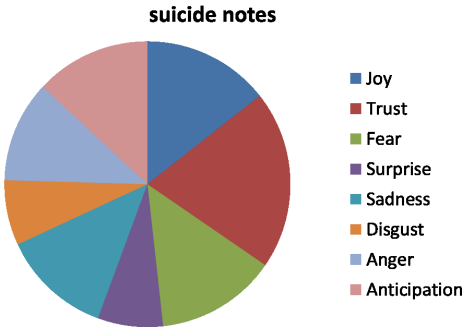

Figure 5: Emotion distribution in suicide notes.

ics and trends.

English-language tweets are then downloaded via the Twitter API using the hashtags as queries. In order to remove both duplicates and highly similar tweets, we used a de-duplication method based on Dice's coefficient. We removed casing, punctuation, hashtags, and URLs, then enforced a maximum overlap of 0.8.

We created our own annotation tool to maximize annotator efficiency and enforce consistency in the annotations. Annotators were provided with an annotation guideline to increase agreement. Annotators were allowed to select any number of emotions for each tweet, or NONE if the tweet had no emotional content. Additionally, HEADLINE, a special case of NONE, was added to have a separate category for tweets containing headlines and links to articles without commentary.

Annotation was split into three phases. In Phase I, the initial teaching phase, three annotators collectively annotated to arrive at a general agreement on an annotation standard. Phase II was an independent annotation phase where 1000 randomly selected tweets were double-annotated. Disagreement was measured (κ = 0.56) and resolved. A smaller set of 500 tweets was then double-annotated and a more reasonable level of agreement (κ = 0.67) was reached. This agreement is somewhat low, but is consistent with emotion annotation on many other tasks (Pestian et al., 2012). Finally, in Phase III, the bulk of the annotating was done individually to maximize the number of annotations. Another 5500 tweets were annotated (for a total of 7000), yielding 500 tweets per topic.

4. Corpus Analysis

The distribution of emotions in the corpus is shown in Figure 2. Note that some tweets contain multiple emotions and are thus over-represented in the graph (20% of tweets have more than one emotion). Furthermore, most tweets have no emotion (57%), so Figure 2 represents only the tweets containing emotion. The most common emotions were DISGUST (16.4% of all tweets) and JOY (12.8% of all tweets), followed by ANGER (10.4%), LOVE (9.2%), and SADNESS (8.8%). Both SURPRISE (5.8%) and FEAR (4.0%) were relatively rare. This distribution is not necessarily representative of all Twitter, however, as they were collected for a few, specific topics.

The distribution of emotions can be compared with those found in love mail (Figure 3), hate mail (Figure 4), and suicide notes (Figure 5). The figures contain a different set of emotions: trust and anticipation are not in our set of emotions, while LOVE is not in the above sets from Mohammad and Yang (2011). Additionally, while the method for determining these emotions used words from emotion lexicons (Mohammad and Yang, 2011) instead of human annotations, some trends can still be distinguished. Notably, our Twitter corpus has significantly more DISGUST than the other data sets. The most similar domain to tweets is the love letters, which contains similar amounts of FEAR, as well as JOY if combined with LOVE. Furthermore, the tweets and love letters have a similar JOY/SADNESS ratio. The primary manner that the tweets diverge from love letters is the amount of ANGER and DISGUST. Actually, there is a significant amount of DISGUST in the tweets, even more so than the hate letters. Again, these ratios are not completely comparable as Mohammad and Yang (2011)

uses an emotion lexicon, but it does reveal that there are likely similarities and differences between these tweets and other domains.

5. Twitter Corpus-Based Emotion Detection

In this section we briefly describe a baseline method for automatically annotating emotions for tweets using the previously described annotated tweets as training data. This baseline is based on the emotion detection method of Roberts and Harabagiu (2012), developed for discovering emotions in suicide notes.

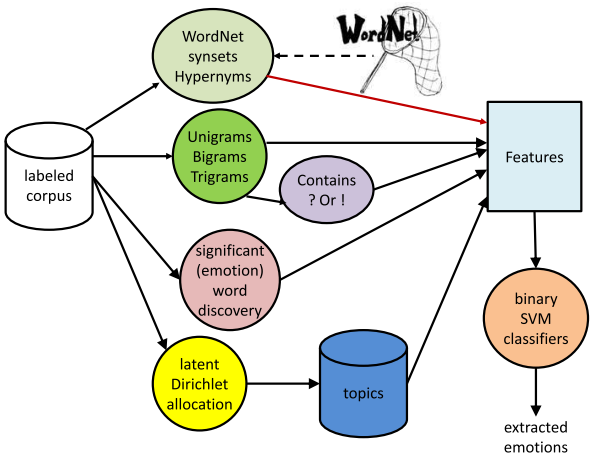

Figure 6: System used for automatically identifying emotions in the corpus.

The system, illustrated in Figure 6, uses a series of binary SVM classifiers to detect each of the seven emotions annotated in the corpus. Each classifier performs independently on a single emotion, resulting in 7 separate binary classifiers implemented using the software available from WEKA (Hall et al., 2009). The combination of these separate classifiers can considered a single multi-label classifier, allowing for a tweet to be annotated with more than one of the emotions (i.e., if multiple binary classifiers return a positive result; if every binary classifier returns a negative result, the sentence has no emotions). Each classifier uses a different set of features, described below. The features used by the binary classifiers are a subset of those employed by (Roberts and Harabagiu, 2012). Notably, the similarity features (which would focus more on topic than emotion) and WordNet Affect features (found to not improve performance) were omitted. One of the differences stems from the usage of WordNet synsets and they possible hypernyms, instead of using the WordNet Affect resource. This decision was made after observing that the tweets do not contain a wealth of the synsets encodes in the WordNet affect. Another difference stems from the usage of bigrams and trigrams instead of phrases. This is due to the existing length constraints of tweets. However, a commonality with the system described in (Roberts and Harabagiu, 2012) consists in the way topics were processed, namely by using modeling techniques, such as latent Dirichlet allocation (LDA). Such techniques can discover similarities between tweets even when tweets have no words in common. We used the MALLET (McCallum, 2002) implementation of LDA and treat every tweet as its own document. LDA then considers every tweet to be associated with a probabilistic mixture of topics, and each topic is composed of a probabilistic mixture of words.

Due to the liberal use of punctuation in tweets, before classification we tokenize on all whitespace and punctuation boundaries, removing URLs, punctuation, and the hash tags used to gather by topic. The features used by the binary classifiers are:

- *Unigrams*: after filtering.
- *Bigrams*
- *Trigrams*
- *Contains !*: A flag indicating the original tweet has an exclamation mark.
- *Contains ?*
- *WordNet synsets*: No word sense disambiguation is performed. Rather, all synsets for each word in WordNet is considered.
- *WordNet hypernyms*: All (recursive) hypernyms for each synset.
- *Topic scores*: The scores for each LDA topic (we use 100 topics).
- *Significant words*: Unigrams judged to have a high pointwise mutual information (PMI) with at least one emotion in the training data. See Section 6. for an explanation of PMI.

For each emotion, the best set of features were chosen with a greedy additive feature selection process. This greedy process iteratively adds the next-best feature to the feature set provided it increases the F_1 score on a development set. Table 3 shows the F_1 scores of our method on each emotion as well as the features used by each emotion classifier. The tests were performed on a 10-fold cross validation, thus allowing tweets from each of the topics to be in both the train and test sets. Interestingly, the best performing emotion was FEAR, which was also the least frequent. Furthermore, the FEAR classifier uses only two features (unigrams and topics). This suggests this emotion is highly lexicalized with less variation than the other emotions, as it has comparable recall but significantly higher precision. The second least frequent emotion, SURPRISE, has the worst performance despite using the second greatest number of features. However, this emotion often involves a great deal of real-world knowledge. For example, given a (bogus) tweet such as "*Napoleon was actually six feet tall*", the only lexical clue is the word *actually*. Otherwise, one would have to know that Napoleon was perceived as being short in order to understand that SURPRISE is being evoked.

6. Linguistic Style of Expressions of Emotion

In "Linguistic Styles: Language Use as Individual Difference", the authors note "*that people differ in the ways they talk and write is hardly a novel observation*" (Pennebaker and King, 1999). Moreover, linguistic *fingerprinting* has often been supported by psychological studies. We extend these observations to the style of expressing emotions in writing as well. One way of measuring linguistic style is provided by the Linguistic Inquiry Word Count

Emotion	#	P	R	F_1	Features
ANGER	583	0.672	0.615	0.642	unigrams, synsets, topics, significant words
DISGUST	922	0.717	0.622	0.666	unigrams, contains !, topics
FEAR	222	0.897	0.629	0.740	bigrams, topics
JOY	716	0.656	0.697	0.676	unigrams, bigrams, contains !, topics
LOVE	516	0.725	0.599	0.656	unigrams, bigrams, trigrams, contains !, contains ?, topics
SADNESS	493	0.747	0.637	0.688	unigrams, contains !
SURPRISE	324	0.631	0.587	0.608	unigrams, contains !, contains ?, topics, significant words
Macro-average	3,777	0.721	0.627	0.668	

Table 3: Emotion detection results for each emotion classifier.

Category	Description	Category	Description
Academ	academic, intellectual, or educational matters	Our	self-inclusive pronouns
AffGain	positive words of love/friendship (e.g., love, date)	Pain	suffering, lack of confidence, or commitment
AffOth	other words of love/friendship (baby, brotherhood)	Persist	endurance
AffPt	affection participant (brother, mother)	Pleasur	enjoyment
ANI	animals	Polit@	political roles (adversary, cabinet)
Aquatic	bodies of water	PowAren	political places
BldgPt	buildings or parts of buildings	PowAuPt	authoritative participants
BodyPt	body parts	PowCon	power conflict (aggression, discord)
COLOR	colors	PowCoop	power cooperation (affiliate, negotiation)
Complet	goal completion	PowDoct	power doctrine (communism, elitism)
Decreas	decrease (cheapen, decay)	PowEnds	goals of the power process
DIST	distance measures	PowPt	power ordinary participants (civilian, follower)
EMOT	emotions	Quality	degrees of quality
EnlEnds	pursuit of enlightenment (contemplate, discover)	Race	racial or ethnic characteristics
EnlLoss	misguided (delude, distract)	RcGain	rectitude gain (worship, forgiveness)
EnlOth	other enlightenment words	RcLoss	rectitude loss (convict, denounce)
EnlPt	enlightenment participant (faculty, historian)	RcRelig	religion (awe, believer)
Exch	buying or selling	Region	general regions (kingdom, downtown)
Exert	exertion	Relig	religious matters (angel, bishop)
Exprsv	arts, sports, or self expression	Rise	rising (ascent, jump)
Fall	falling (sing, tumble)	Ritual	social rituals (baseball, birthday)
Feel	feelings (gratitude, apathy)	Role	social roles (actor, colleague)
Female	women and their social roles	RspLoss	losing of respect
Food	food and beverage	Say	say and tell
FREQ	frequency or recurrence	SklAsth	skill aesthetic (beautiful, poetic)
Goal	end states for mental or physical effort	SklOth	other skill words (adept, blunder)
Intrj	interjections	SklPt	skill participant (baker, carpenter)
IPadj	relations between people (unkind, aloof)	Sky	aerial or outer-space conditions (haze, rain, sun)
Kin@	kinship	Think	rational thought process
Know	awareness, certainty, similarity and antonyms	TIME	temporal (afternoon, decade)
Land	natural places (desert, beach)	Vehicle	vehicle (jet, limousine)
MALE	men and their social roles	WlbGain	gain in well being (comfort, feed)
Milit	military matters	WlbPhys	physical aspects of well being (bone, cancer)
Name	demonyms (Cuban, African)	WlbPsyc	psychological aspects of well being (anger, cry)
Nation	country names and demonyms	WlbPt	well being participant (nurse, baby)
Nonadlt	infants/adolescents	WlbTot	all well being words
ORD	ordinal words	WltTran	wealth transaction (import, mortgage)
Ought	moral imperative	You	pronouns for another person

Table 4: General Inquirer semantic categories that were significant for certain emotions and/or topics in our corpus.

(LIWC) method (Pennebaker and King, 1999), which measures word use in psychologically meaningful categories that capture attentional focus (through pronouns and verb tenses), emotionality (words that express positive or negative emotions), words that signal social relationship, social coordination, status and social hierarchies, as well as words that indicate honesty and deception. Danescu-Niculescu-Mizil et al. (2011) employs 16 of the 60 categories of words for modeling style accommodation in Twitter communications. No emotion categories were considered in that study.

As LIWC contains few categories to describe emotional content, we have searched for additional psycholinguistic resources that may provide insights into the style of conveying emotions on Twitter. In considering using the LIWC or a similar approach for analyzing the Twitter styles that convey emotions, we took note of the observation that although we typically use a vocabulary of almost 100,000 words for composing tweets, only about 500 of them are style words, which are typically function words (e.g., *the*, *but*, *without*). We believe it is quite difficult to correlate

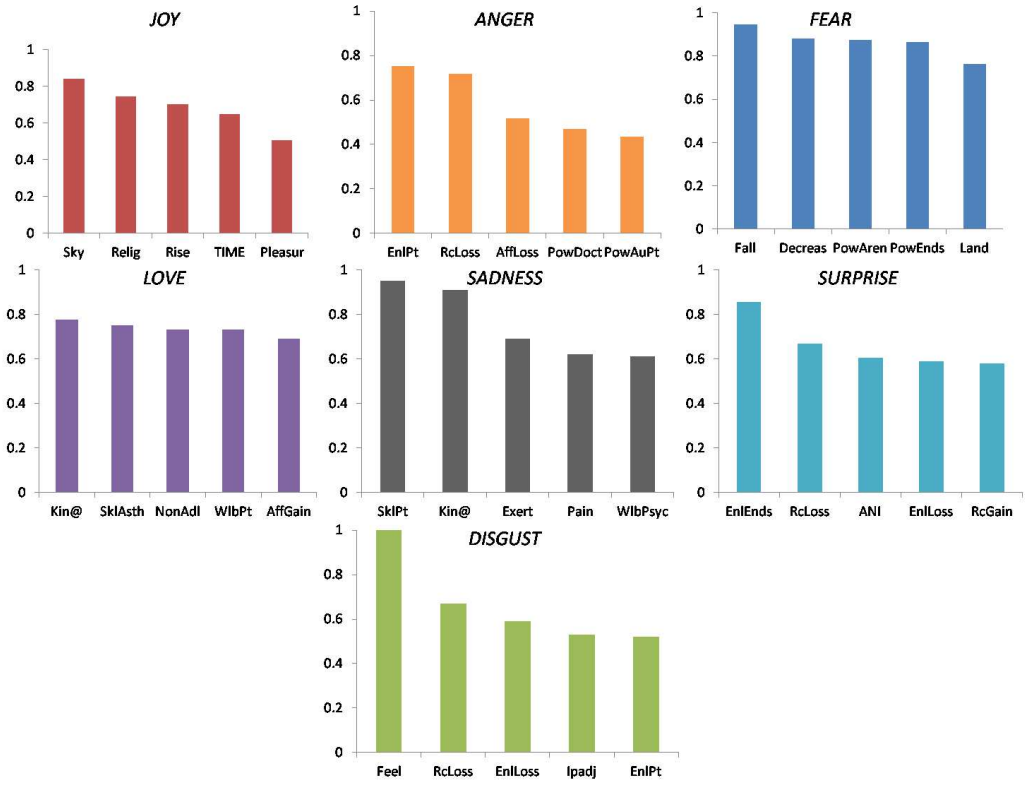

Figure 7: Most significant General Inquirer categories for each emotion from our corpus.

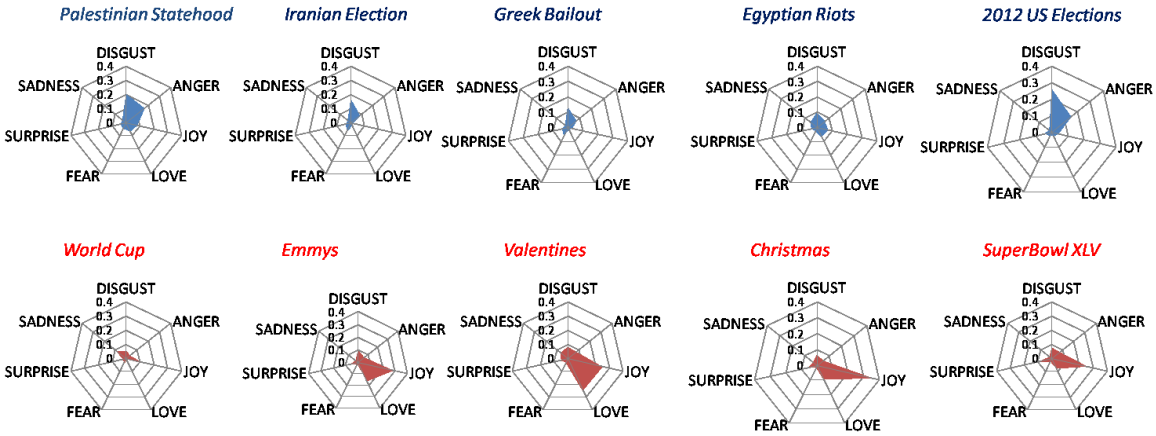

Figure 8: Emotion distributions for ten Twitter topics.

only style words and their patterns to emotions, especially as style words make up about 55% of the words we speak, hear, or read, but only 20% of the words in our social media interactions. In consequence, we decided to make use of one of the first general-purpose computerized text analysis resources developed in psychology, namely the General Inquirer (Stone et al., 1966).

Originally, the General Inquirer technique relied on the Harvard psychological dictionaries that were correlated with states, motives, social and cultural roles as well as different aspects of general distress. The current version of the General Inquirer also contains lexical categories, which we ignored, and hundreds of different semantic categories, some of which are listed in Table 4. We are interested in learning: (1) which semantic categories best define the style of each of the emotions encoded in our corpus, (2) which semantic categories are characteristic of each of the topics we have considered when building the corpus, and (3) the distribution of emotions for each topic. In measuring (1) and (2), we must avoid categories that are prevalent due to their high frequency. We thus employ the pointwise mutual information (PMI) metric (also used in the features above):

$$\text{PMI}(x, c) = \log \frac{p(x, c)}{p(x)p(c)} = \log \frac{p(c|x)}{p(c)}$$

Where c is a category from the General Inquirer and x is either an emotion or topic.

Figure 7 illustrates the most representative semantic categories from the General Inquirer that were discovered for each of the seven emotions annotated in our corpus.

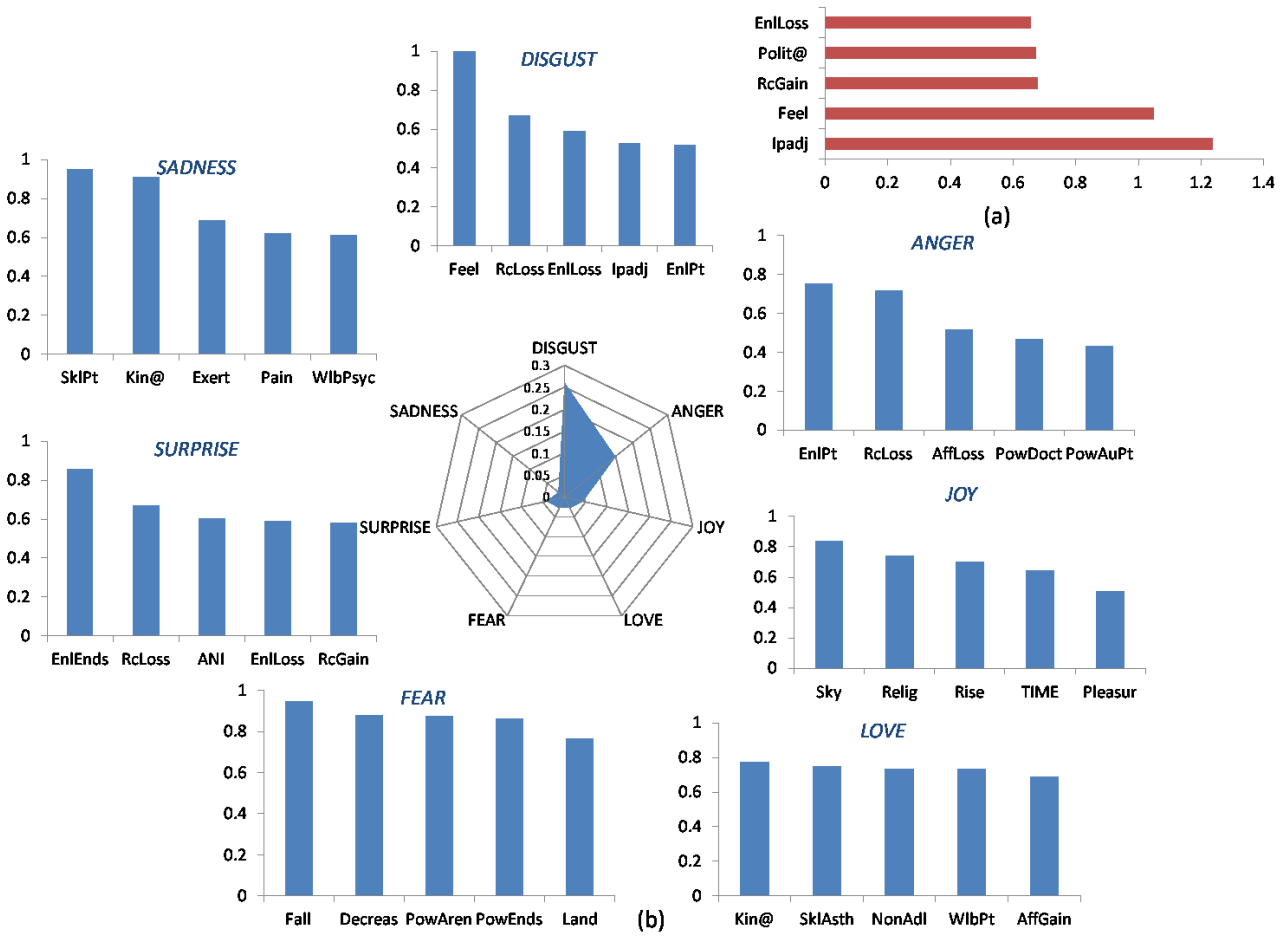

Figure 9: (a) Linguistic style features for the topic "2012 U.S. Election". (b) Emotion make-up and corresponding linguistic style semantic categories for the topic "2012 U.S. Election".

Surprisingly, JOY is commonly expressed using terms for aerial conditions (haze, rain, sun, etc.) despite the fact that none of our topics are highly associated with the sky. Other common categories when JOY is evoked are religion, largely due to thankfulness, and rising, largely metonymic terms to suggest improvement (e.g., rise, soar, leap). These semantic categories were more strongly associated with JOY than semantic categories that specifically target enjoyment. Figure 7 also illustrates that in the case of LOVE, in almost equally high degree, words about kinship, skills and aesthetic, as well as words about children, are more indicative than words about affection or friendship. In the case of the emotion FEAR, the linguistic style of tweets, as captured by the General Inquirer, evokes semantic categories regarding falling, decreasing, words about political places, or indications of goals of obtaining power, and, surprisingly, words about natural places (such as a beach or desert). In the case of the remaining emotions, there seems to be one or two dominant categories for each: enlightenment participant and a moral sinner (such as a convict) for ANGER; words of feeling in the case of DISGUST; words of pursuit of enlightenment seem to stylistically better express SURPRISE; whereas SADNESS is mostly expressed in tweets by words that describe occupations (e.g., baker, carpenter) or kinship, commonly to describe a lost loved one.

Our analysis revealed that the stylistic fingerprinting of each of the topics was influenced by the distribution of emotions evoked in their respective tweets. Figure 8 illustrates the emotional make-up of ten of the topics from our corpus. It is interesting to note that many topics that could be considered to have similar themes (e.g., topics related to politics) have different emotional make-ups, suggesting differing views on each subject. The Palestinian Statehood and the U.S. elections topics stand out due to the intensity of DISGUST and ANGER. In the Egyptian riots topic, DISGUST is less intense, but it is mixed with JOY and LOVE. In the Iranian Election topic, FEAR also plays a role, however not as important as ANGER. Dominated by DISGUST, the Greek bailout topic also combines FEAR and SURPRISE. The emotional make-up of the other topics illustrated in Figure 8 are dominated by JOY. In the case of the topic of Valentines Day, it is combined with an intense feeling of LOVE, much less with SURPRISE and SADNESS. JOY is also the predominant emotion expressed in the Superbowl topic, but with much less degree of SURPRISE or DISGUST. Figure 8 illustrates how each topic is characterized by a different statistical make-up of emotions. Thus we are also interested in analyzing the linguistic style associated with each topic and also its style of expressing emotions.

Each topic is characterized by the style in which informa-

tion is expressed in tweets. Figure 9(a) illustrates the more representative semantic categories for the topic of the 2012 U.S. Election, showing that words that indicate relations between people dominate the style of those tweets, as well as words of feeling. However, knowledge about the emotional make-up of the tweets in a topic are expected to contribute to the understanding of the linguistic style used for expressing those emotions. That linguistic style is completely different than the style that characterizes the topic in general. Figure 9(b) illustrates both the emotional make-up for the 2012 U.S. Election and the semantic categories from the General Inquirer associated with the emotions. DISGUST is the most representative emotion, expressed through words of feeling. Words of feeling are also representative for the style of the tweets in the topic, but words about the relations between people (IPadj) contribute less to the anticipation of emotions than words about rectitude loss (RcLoss). What is surprising is that since the EnlPt category (describing enlightenment participants) contributes as much to expressions of DISGUST as words from the IPadj category, they do not seem to characterize the topic as much as words of political roles (Polit@). These observations lead us to believe that perhaps a better discrimination of emotions in a corpus can be achieved by taking into account the differences between the topic-specific linguistic style and the emotion-specific linguistic style. We plan to use these observations for designing both supervised and unsupervised methods for automatic emotion detection that take advantage of these style differences.

7. Conclusion

We have described the creation of an emotion corpus created from the micro-blogging service Twitter. The corpus contains seven different emotions annotated across 14 topics. We have developed a baseline approach to use for benchmarking and used this approach to compare our corpus with several existing emotion corpora. We have also shown that a simple supervised method for detecting emotions can be trained on this corpus. Moreover, we have conducted an analysis of the emotional make-up of the topics that make up the corpus and have characterized the linguistic style of each topic and each emotion in the corpus. This analysis should lead to the design of novel supervised and unsupervised emotion detection techniques.

8. References

Johan Bollen, Huina Mao, and Alberto Pepe. 2011a. Modeling Public Mood and Emotion: Twitter Sentiment and Socio-Economic Phenomena. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, pages 450–453.

Johan Bollen, Muina Mao, and Xiaojun Zeng. 2011b. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8.

Cristian Danescu-Niculescu-Mizil, Michael Gamon, and Susan Dumais. 2011. Mark my words! Linguistic style accommodation in social media. In *World Wide Web*.

Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. The WEKA Data Mining Software: An Update. *SIGKDD Explorations*, 11(1).

Sepandar D. Kamvar and Jonathan Harris. 2011. We Feel Fine and Searching the Emotional Web. In *Fourth ACM International Conference on Web Search and Data Mining*.

Suin Kim, JinYeong Bak, Yohan Jo, and Alice Oh. 2011. Do You Feel What I Feel? Social Aspects of Emotions in Twitter Conversations.

Andrew Kachites McCallum. 2002. Mallet: A machine learning for language toolkit.

Saif M. Mohammad and Tony Yang. 2011. Tracking Sentiment in Mail: How Genders Differ on Emotional Axes. In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA)*, pages 70–79.

Saif M. Mohammad. 2011. From Once Upon a Time to Happily Ever After: Tracking Emotions in Novels and Fairy Tales. In *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage*, pages 105–114.

Alexander Pak and Patrick Paroubek. 2010. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation*.

Bo Pang and Lillian Lee. 2008. Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, 2(1–2):1–135.

James W. Pennebaker and Kaura A. King. 1999. Linguistic Styles: Language Use as Individiual Different. *Journal of Personality and Social Psychology*, 77(6).

John P. Pestian, Pawel Matykiewicz, Michelle Linn-Gust, Jan Wiebe, Kevin Cohen, Christopher Brew, John Hurdle, Ozlem Uzuner, and Brett South. 2011. Sentiment Analysis of Suicide Notes: A Shared Task (Submitted). *Biomedical Informatics Insights*.

John P. Pestian, Pawel Matykiewicz, Michelle Linn-Gust, Brett South, Ozlem Uzuner, Jan Wiebe, Kevin B. Cohen, John Hurdle, and Christopher Brew. 2012. Sentiment Analysis of Suicide Notes: A Shared Task. *Biomedical Informatics Insights*, 2012(5 (Suppl. 1)).

Kirk Roberts and Sanda Harabagiu. 2012. Statistical and Similarity Methods for Classifying Emotion in Suicide Notes. *Biomedical Informatics Insights*, 2012(5 (Suppl. 1)).

Philip J. Stone, Dexter C. Dunphy, and Marhsall S. Smith. 1966. *The General Inquirer: A Computer Approach to Content Analysis*. MIT Press.

Carlo Strapparava and Rada Mihalcea. 2008. Learning to Identify Emotions in Text. In *Proceedings of the ACM Conference on Applied Computing*.