EMOTEX: Detecting Emotions in Twitter Messages

Maryam Hasan, Elke Rundensteiner, Emmanuel Agu
Computer Science Department, Worcester Polytechnic Institute
mhasan@wpi.edu, rundenst@cs.wpi.edu, emmanuel@cs.wpi.edu

Abstract

Social media and microblog tools are increasingly used by individuals to express their feelings and opinions in the form of short text messages. Detecting emotions in text has a wide range of applications including identifying anxiety or depression of individuals and measuring well-being or public mood of a community. In this paper, we propose a new approach for automatically classifying text messages of individuals to infer their emotional states. To model emotional states, we utilize the well-established Circumplex model that characterizes affective experience along two dimensions: valence and arousal. We select Twitter messages as input data set, as they provide a very large, diverse and freely avail- able ensemble of emotions. Using hash-tags as labels, our methodology trains supervised classifiers to detect multiple classes of emotion on potentially huge data sets with no manual effort. We investigate the utility of several features for emotion detection, including unigrams, emoticons, negations and punctuations. To tackle the problem of sparse and high dimensional feature vectors of messages, we utilize a lexicon of emotions. We have compared the accuracy of several machine learning algorithms, including SVM, KNN, Decision Tree, and Naive Bayes for classifying Twitter messages. Our technique has an accuracy of over 90%, while demonstrating robustness across learning algorithms.

1 Introduction

1.1 Background

Social networks and microblogging tools such as Twitter allow individuals to express their opinions, feelings, and thoughts on a variety of topics in the form of short text messages. These short messages (commonly known as tweets) may also include the emotional states of individuals (such as happiness, anxiety, and depression) as well as the emotions of a larger group (such as opinions of people in a certain country or affiliation) [1, 2]. In fact, Twitter can be considered a large repository that includes a rich ensemble of emotions, sentiments and moods. For example the tweet "Great Christmas spent with my amazing family" expresses a happy mood and the tweet "Feelings Hurt Tonight!" expresses sadness. Table 1 provides more examples of tweets expressing different types of emotion.

Emotion	Tweet
Happy	Excited to see him in Texas in two weeks. So many weddings coming up how exciting is that.
Relaxed	I feel so at peace right now. The sound of rain always puts me to sleep.
Stressed	Seriously stressed over this final. Presentation? I'm feeling like I'm waiting to get an injection!
Depressed	RIP Grandpa, you will be missed. I'm just so #depressed and on the verge of crying.

Table 1: Examples of Emotion Tweets

1.2 Motivation

In this paper, we investigate a method for automatically detecting and classifying the emotions expressed by Twitter messages. A system developed based on this method could potentially be employed in a large variety of applications, ranging from well-being apps, self-helps, counselors, to community population studies.

The proposed method can be used by healthcare professionals or counseling agencies to monitor and track a patient's emotional states, or to recognize anxiety or systemic stressors of populations (e.g. different student groups on campus). The system can also help commercial agencies to gauge sentiment of buyers or to facilitate targeted product advertisement.

In addition, this technology can measure public mood of people in a community, which may help social scientists to understand the quality of life of populations. Measuring and tracking the living conditions and quality of life of a society are essential for public policy making. The quality of life can be measured based on different aspects of life including social, emotional, psychological, life satisfaction, and work. However, methods that measure living conditions fail to measure what people think and feel about their lives, such as their positive or negative emotions, or their overall satisfaction with life [3, 4]. The quality of life is typically measured using self-reports and surveys [5]. People are asked to fill out questionnaires about their life and their day-to-day emotions. Collecting these questionnaires is very time consuming, tedious, and error-prone. However, the system developed based on our proposed approach would

 ISBN: 978-1-62561-000-3

be able to automatically detect what people feel about their lives from twitter messages. For example, the system can recognize:

- percentage of people expressing higher levels of life satisfaction in one group versus another group,
- percentage of people who feel happy and cheerful,
- percentage of people who feel calm and peaceful, and
- percentage of people expressing higher levels of anxiety or depression.

Classifying text messages based on their emotion or sentiment is a growing area of research. However, although some prior work has been done to classify Twitter messages (see related work in Section 5), most of them have focused on determining message sentiment instead of emotion [6, 7, 8, 9]. Sentiment refers to the opinions of individuals about a topic (e.g. a movie or a new product). Sentiment is categorized either as positive or negative. Our goal instead is to provide an approach for automatically and accurately classifying Twitter messages into distinct emotional categories. To represent classes of emotion, we adopt the Circumplex model [10], a popular model of human emotions, which characterizes affective experience through two dimensions: valence and arousal. Instead of classifying tweets into two classes (positive or negative) as in sentiment analysis, we design methods to detect and classify short text messages into four finer-grained classes of emotion.

1.3 Challenges

In order to classify texts into emotional categories accurately, several challenges should be tackled:

Casual Twitter language and noise: Tweets are casual, contain numerous punctuation and spelling errors and are limited to 140 characters of text. While the use of informal language and short messages have been previously studied in the context of sentiment analysis [7, 8, 11, 12], the use of such language to express emotions has been much less studied.

Large numbers of potential features: The large number of features are available to categorize short text messages. Each tweet is presented as a vector of features which is an n-dimensional vector of numerical values. Single words that exist in the input dataset are potential features to be included in the feature vector. With the large breadth of topics discussed on Twitter, the number of words in our dictionary becomes very large. As a result the feature vector for each tweet will be very large and sparse (i.e. many features will have a value of zero).

Labeling for supervised learning: Text messages, in their raw form, do not have labels. However, in order to train a classifier, supervised learning methods require labeled data. With the large volumes of Twitter messages, it would be time consuming and tedious to manually label tweets and then train a classifier for emotions.

Crowdsourcing is one popular approach for labeling data [13, 14, 15]. Tools such as Amazon's mechanical Turk provide easy access to large numbers of manual data labelers and annotators. However, annotators may not always be reliable. Their interpretations of the same data may be ambiguous even when clear instructions are given. For emotion classification, using humans as annotators does not guarantee that they can correctly infer the author's emotional state.

Researchers have started to investigate rules and automatic methods for labeling training data. For example, Go et al. [7] and Pak and Paroubek [8] used Western-style emoticons as labels to classify Twitter messages as having either positive or negative sentiment [7, 8]. Barbosa and Feng [11] used existing Twitter sentiment sites for collecting training data, they also used 1000 manually labeled tweets for tuning and another 1000 manually labeled tweets for testing.

1.4 EMOTEX: Our Proposed Approach

Our proposed approach resolves the challenges mentioned in previous section as following:

To support casual Twitter language and noise: correct misspellings and casual language by pre-processing all Tweets with clearly defined rules.

To label Tweets: use Twitter hash-tags as labels that indicate the emotion expressed by Tweets. For example, a tweet with the hash-tag "#depressed" is labeled as one expressing a depressed emotion, while a tweet containing the hash-tag "#excited" is labeled as expressing excitement. Using the Twitter API, we collected a large number of tweets with hash-tags that then served as noisy labeled data. While hash-tags can themselves cause errors or be ambiguous when utilized as labels, they enable the automatic processing of tweets that would otherwise take hours or days to label manually. After extracting enough labeled data, the hash-tags were removed in order to force the classifier algorithm to learn from other features.

To avoid high dimensional and sparse feature vectors: use a lexicon of emotions. Our dictionary does not include all words in the training dataset, but instead it focuses on the emotional words from the lexicon LIWC (Linguistic Inquiry & Word Count: http://www.liwc.net/) [16]. The LIWC contains several thousand words. We use emotion-indicative categories such as positive emotions, negative emotions, anxiety, anger and sadness to build our domain-specific dictionary.

In particular, EMOTEX makes the following contributions:

- Designing and implementing a method to automatically label twitter messages according to the emotions of their authors.
- Resolving the problem of high dimensional feature space in twitter dataset.
- Achieving highest accuracy for classifying twitter messages based on their emotional states.

The rest of the paper is organized as follows. Section 2 describes the model that we exploit to categorize emotional states of individuals. In section 2 we describe our proposed approach in Emotex. In section 4 we present our experiments and discuss the results. In Section 5 we discuss prior work on sentiment analysis and emotion analysis on micro-blog data. We conclude and give future directions of research in section 6.

2 Model of Emotions

The categorization of emotions has mainly been studied from two fundamental approaches: basic emotions and core affect.

2.1 Model of Basic Emotions

Basic emotion theorists believe that humans have a small set of basic emotions, which are discrete.

Various researchers have attempted to identify a number of basic emotions which are universal among all people and differ one from another in important ways. A popular example is a cross-cultural study of 1972 by Paul Ekman and his colleagues, in which they concluded that the six basic emotions are anger, disgust, fear, happiness, sadness, and surprise [17].

Consequently most work in the field of emotion mining and classification from text has been based on this basic emotion sets [1, 18, 19]. For example, in order to model public mood and emotion, Bollen *et al* extracted six dimensions of mood including tension, depression, anger, vigour, fatigue, confusion from Twitter [1]. Strapparava and Mihalcea annotated a large data set with six basic emotions: anger, disgust, fear, joy, sadness and surprise [19].

However, there is no consensus amongst theorists on which human emotions should be included in the basic set. Moreover, the distinction of one emotion from another is a contested issue in emotion research. For instance, it is unclear if "surprise" should be considered an emotion since it can assume negative, neutral or positive valence.

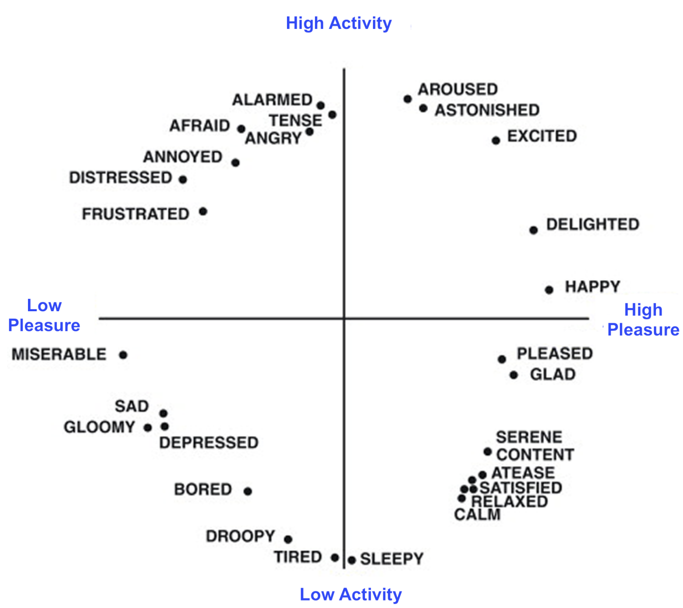

Figure 1: Circumplex Model of Affect including 28 affect words by J. A. Russell, 1980. Used with permission. [10]

2.2 Core Affect Model

Despite of the basic emotions model which defines discrete emotions, the core affect model defines emotion on a continuous scale [20]. Core affect model of emotion characterizes human emotions by defining their positions along two or three dimensions. That is, most dimensional models incorporate valence and arousal dimensions.

One of the very first practical models of core affect is Russell's Circumplex Model of Affect [10]. As shown in Figure 1, the model suggests that emotions are distributed in a two-dimensional circular space, containing pleasure and activation dimensions. The activation dimension measures if one is likely to take an action in a mood state. The pleasure dimension measures how positive or negative a person feels. The vertical axis represents activation or arousal, and horizontal axis represents pleasure or valence. The center of the circle represents a neutral valence and a medium level of arousal.

The Circumplex model has been well validated and widely used in other studies [14]. We utilize the Circumplex model by considering four major classes of emotions: Happy-Active, Happy-Inactive, Unhappy-Active, and Unhappy-Inactive. This model is simple, and describes a wide range of emotional states we have selected for our work. Moreover the four classes of emotions are very distinct, because each class constitutes emotions which are quite different compared with the emotions of other classes.

3 EMOTEX: Detecting Emotions in Text Messages

To detect emotions in text messages such as tweets, we apply supervised learning methods to automatically classify short texts, according to a finer-grained category of the emotions.

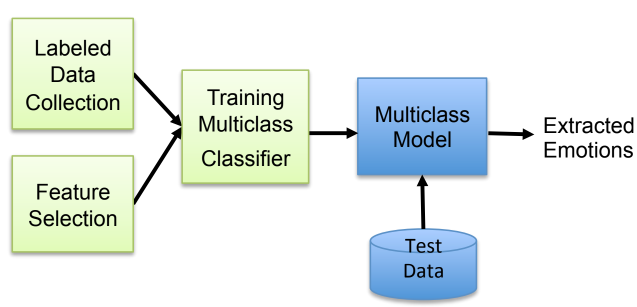

Figure 2: Model of Emotex

This section describes in more detail how Twitter messages are collected, labeled and classified according to the emotions they convey. Figure 2 shows the process flow of Emotex. We gathered Tweeter messages, selected features and trained classifiers that classify tweets into multiple emotion classes.

 ISBN: 978-1-62561-000-3

3.1 Collecting Labeled Data

Twitter message features such as hash-tags and emoticons are likely to be useful features for sentiment and emotion classification.

The usage of hashtags in tweets is very common, and Twitter dataset contains millions of different user-defined hash-tags. A study of a sample of 0.6 million tweets by Wang *et al.* [21] showed that 14.6% of tweets in their sample had at least one hashtag. Many tweets include more than a single tag. These hash-tags could help to group messages that indicate a certain emotion. Our results (see Section 4) confirms that hash-tags are indeed useful features for sentiment and emotion classification.

In order to collect labeled data, we need to identify the list emotion hash-tags. First, we exploit the set of 28 affect words from Circumplex model (as shown in Figure1) as initial set of keywords. Then we extend the initial keywords using WordNet's synsets, and we use them to find emotion hash-tags. Using the list of emotion hash-tags, we collect tweets which contain emotion tags. Twitter has an API that can be used to automatically collect tweets and filter them by query terms or hash-tags. Figure 3 shows the steps of data collection.

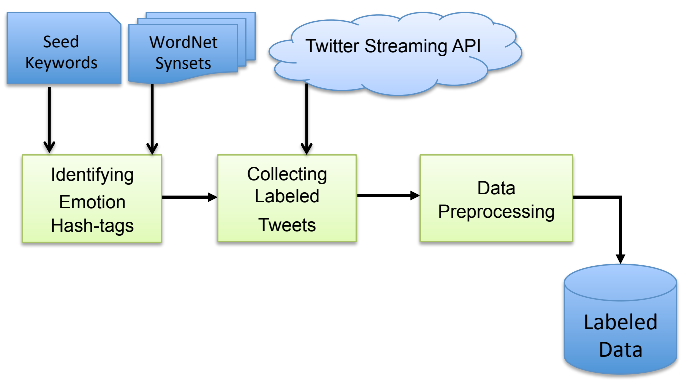

Figure 3: Model of labeled data collection

3.1.1 Preprocessing

After collecting the same number of tweets for each class, the labeled tweets are then pre-processed to mitigate misspellings and casual language used in Twitter using the following rules:

1. Tweets often contain usernames which start with the @ symbol before the username (e.g. @Marilyn). All words that start with the @ symbol are replaced by "USERID".

2. Many tweets contain url links. All the url links are replaced with the "URL".

3. Words with repeated letters such as happyyyyy, are common in Twitter messages. Any letter occurring more than two times consecutively is replaced with one occurrence. For instance, the word "happyyyyy" would be changed into "happy".

4. Many tweets contain more than one hash-tag, while some may even contain hash-tags from two different classes. For example the tweet "Got a job interview today with At&t... #nervous #excited.", includes hash-tag #nervous from Unhappy-Active class and tag #excited from Happy-Active class. Any tweet containing hash-tags from different classes are removed from training data. Tweet may also be removed if they contain two subjects. Such tweets are removed because they would introduce ambiguities into our training set. We do not want features of one class marked as part of another class. These tweets correspond to a mixture of emotion of different classes; therefore they would mislead our classifier algorithm.

5. Some tweets contain emoticons from two different classes. For example the tweet "Tomorrow, first volleyball match :) and final exam :(", includes both happy and sad emoticons. These tweets are removed during pre-processing.

6. Some tweets contain conflicts between hash-tags and emoticons. For example the tweet "Yup, I'm totally considering leaving this planet now :) #disappointed #nohope", includes hash-tag "#disappointed" from Unhappy-Inctive class and emoticon ":)" which shows happiness. The tweets from unhappy-Inactive class containing happy emoticons as well as tweets from happy-active class containing sad emoticons are removed from labeled data.

7. In Twitter, hash-tags can be placed in the beginning, middle, or end of a tweet. As part of pre-processing, hash-tags are stripped off from the end of tweets. For instance, the tags "#disappointed" and "#sad" are removed from the tweet "No one wants to turn up today. #disappointed #sad ". If the tags were not stripped off, then the classifiers tend to put a large amount of weight on the tags, which may hurt accuracy. However the tags in the beginning or in the middle of the tweet are left, since they are part of the content of the sentence. For example in the tweet "That #nervous but hopeful #feeling that keeps you up at night or makes you get up early" the tags "#nervous" and "#feeling" are part of the sentence and are kept.

3.2 Feature Selection

In order to train a classifier from labeled data, we represent each tweet into a vector of features. We need to capture features that describe the emotion expressed by each tweet. Feature selection plays an important part in the effectiveness of the classification process. For this study, we explore the usage of different features. We use single words, also known as unigrams as the baseline features for comparison. Other features explored included the presence of emoticons, punctuations, and negations, as elaborated below.

3.2.1 Unigram Features

Unigrams or single word features have been widely used to capture the sentiment or emotion of a tweet[7, 8, 18].

ISBN: 978-1-62561-000-3

Let $(f_1, f_2, ..., f_m)$ be our predefined set of unigrams that can appear in a tweet. Each feature f_i in this vector is a word from the dictionary of words in our dataset. Text messages can be classified into emotion categories based on the presence of affect words like "annoyed", and "happy". Therefore, the problem of high dimensional feature vector can be solved by identifying an appropriate emotion lexicon. We effectively design a domain-specific dictionary by using the lexicon of emotions, instead of all the words in our input dataset.

As a result, our feature space does no longer include all the words in our training dataset, but instead it only contains the emotional words from the emotion lexicons. This method reduces the size of feature space dramatically, without loosing informative terms. We decide to use the LIWC lexicon, which has been well validated and widely used in other studies [22, 23, 15]. LIWC contains a dictionary of several thousand words, wherein we use emotion-indicative categories such as positive emotions, negative emotions, anxiety, anger, sadness, and negation and utilize them effectively as our domain-specific dictionary.

3.2.2 Emoticon Features

Other than unigrams, emoticons are likely to be useful features for emotion classification in text messages since they are textual portrayals of a writer's emotion in the form of icons. These features tend to be widely used in sentiment analysis. Go *et al* and A. Pak *et al* [7, 8] utilized the western-style emoticons to collect labeled data. There are many emoticons that can express happy emotion, sad emotion, annoyed emotion or sleepy emotion. For example, ":)" and ":-)" both express happy emotion. The full list of emoticons that we used can be found in Figure 4.

3.2.3 Punctuation Features

Other features potentially helpful for emotion detection are punctuations. Users often use exclamation marks when they want to express their strong feelings. For instance, the tweet "I lost 4lb in 3 days!!!!!" expresses strong happiness and the tweet "we're in december, which means... ONE MONTH UNTIL EXAMS!!!" represents a high level of stress. The exclamation mark is sometimes used in conjunction with the question mark, which indicates astonishment. For example the tweet "You don't even offer high speed anymore, yet you keep overcharging me?!" indicates an astonished and annoyed feeling.

3.2.4 Negation Features

As our last feature, we select negation to address errors caused by tweets that contain negated phrases like "not sad" or "not happy". For example the tweet, "It's Christmas and I've gota get up in 6h, to get a plane back to shitty England!! Not happy" should be classified as a sad tweet, even though it has a happy unigram. To tackle this problem we define negation as a separate feature. We selected the list of negated phrases from the LIWC dictionary.

In summary, we investigate four feature types for emotion classification: single word features, emoticon features, punctuation features, and negation features. For the classification, all feature types are combined into a single feature vector.

3.3 Classifier Selection

A number of statistical classification techniques have been applied to text categorization, including regression models, Bayesian classifiers, decision trees, nearest neighbor classifiers, neural networks, and support vector machines. For the task of classification we used four different classifiers including, support vector machine, Naive Bayes, Decision Trees, and K-Nearest Neighbors, which have been shown to be effective in previous text classification work.

SVM Classifiers attempt to partition the data space using linear or non-linear boundaries between different classes. SVMs achieve high performance in text categorization since they accept high dimensional feature spaces and sparse feature vectors. Also text classification using SVMs is very robust to outliers and does not require any parameter tuning [24].

Bayesian classifiers build a probabilistic model based on the word features in different classes. Texts are classified based on posterior probabilities generated based on the presence of different classes of words in texts. Naive Bayes has been widely used for classifying text because it is simple and fast.

KNN classifies new text by a majority vote of its neighbors based on a similarity measure (e.g., distance functions). The KNN algorithm is fast and simple, but it is sensitive to the local structure of the data.

In the Decision Tree classifier, leaves are class labels and branches represent conjunctions of features that lead to those class labels. Decision trees are slow and sometimes suffer from over-fitting. However, its accuracy competes with well known text classification algorithms such as SVM.

Category	Emoticons
Happy Emoticons	:) ;) =) :] :P :-P ;P :D ;D :> :3 :-) ;-) :^) :o) :~) ;^) ;o) :') :-D :->
Sad Emoticons	:(=(:-(:^(:o(:^(:'(:-<
Angry Emoticons	>:S >:{ >: x-@ :@ :-@ :-/ :-\ :/
Afraid/Surprised Emoticons	:-o :-O o_O O_o :$
Sleepy Emoticons	-_- ~_~

Figure 4: List of Emoticons

4 Experimental Results

4.1 Identifying Emotion Hash-tags

In order to collect labeled data, we identified the list of hash-tags corresponding to each class of emotions. As we

 ISBN: 978-1-62561-000-3

mentioned in Section 2 we selected four classes of emotions namely Happy-Active, Happy-Inactive, Unhappy-Active, and Unhappy-Inactive. First we selected an initial set of keywords for each category from the Circumplex model as shown in Figure 1. We only selected those keywords which are very distinct and distinguishable from other classes. We ignored the keywords that are located close to the boundary of four dimensions. The initial list of keywords are listed below:

- Class Happy-Active: Happy, Excited, Delighted, Astonished, Aroused
- Class Happy-Inactive: Serene, Contented, Satisfied, Relaxed, Calm
- Class Unhappy-Active: Tense, Angry, Afraid, Annoyed, Distressed
- Class Unhappy-Inactive: miserable, Depressed, Sad, Gloomy

Class	Hash-tags
Happy-Active	#elated,#overjoyed,#enjoy,#excited, #proud,#joyful,#feelhappy,#sohappy, #veryhappy,#happy,#superhappy, #happytweet,#feelblessed,#blessed, #amazing,#wonderful,#excelent, #delighted,#enthusiastic **example: Thankful for unexpected time with one of my best friends #happy**
Happy-Inactive	#calm,#calming,#peaceful,#quiet, #silent,#serene,#convinced,#consent, #contented,#contentment,#satisfied, #relax,#relaxed,#relaxing,#sleepy, #sleepyhead,#asleep,#resting, #restful,#placid **example: ready for a relaxing day of doing nothing #relaxing**
Unhappy-Active	#nervous,#anxious,#tension,#afraid, #fearful,#angry,#annoyed,#annoying, #stress,#distressed,#distress,#stressful, #stressed,#worried,#tense,#bothered, #disturbed,#irritated,#mad,#furious **example: I have my speech in less than minutes #nervous**
Unhappy-Inactive	#sad,#ifeelsad,#feelsad,#sosad, #verysad,#sorrow,#disappointed, #supersad,#miserable,#hopeless, #depress,#depressed,#depression, #fatigued,#gloomy,#nothappy, #unhappy,#suicidal,#downhearted, #hapless,#dispirited **example: Sometimes people let you down and it hurts. #sad**

Table 2: List of hash-tags for each emotion class

After defining the initial set of keywords for each category, we extended the list by using set of synonyms defined by WordNet. We also added hash-tags to our list from Twitter, namely emotion-specific tags such as the tag "#ifeelsad". Using the extended list of keywords, we obtained a set of 20 unique emotion hash-tags for each class. The tags of each class constitute emotions which are quite different compared with the emotions of the other classes. Table 2 presents the final list of hash-tags used for collecting labeled data for each class.

Table 3 represents the number of collected labeled tweets before and after pre-processing. As it shows the number of tweets decreased by 19% by removing noisy tweets during preprocessing.

Class	Number of Tweets Before Pre-processing	Number of Tweets After Pre-processing
Happy-Active	39600	34000
Happy-Inactive	41000	29200
Unhappy-Active	44000	37000
Unhappy-Inactive	40700	33900
Total	165300	134100

Table 3: Number of Tweets collected as labeled data

The histogram in Figure 5 represents the distribution of four classes of tweets that we collected and labeled using hash-tags, during the new year vacation and after it. It shows that the number of happy tweets during the vacation are higher than the number of happy tweets after vacation by 12%, as we expected. However the number of happy tweets didn't change significantly (only 1%) between one week after new year and two weeks after it.

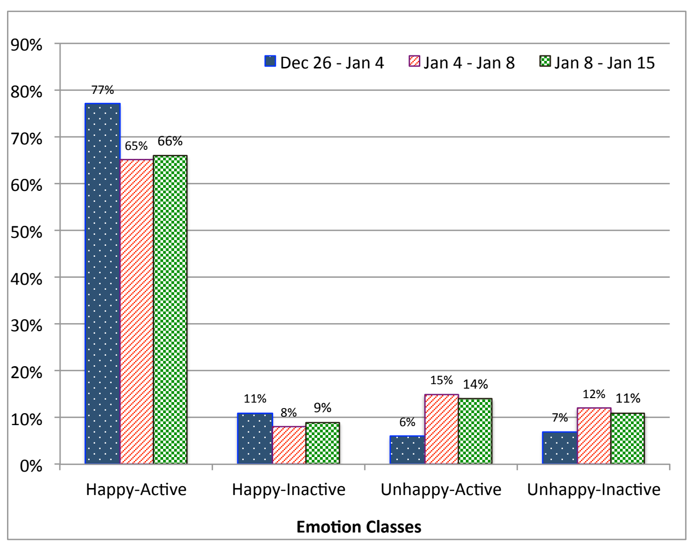

Figure 5: Distribution of the emotions in collected data during new year vacation and after it

After identifying hash-tags, we collected labeled data for each class. Twitter has an API that can be used to automatically collect Tweets by query terms or hash-tags. We

Features	Naive Bayes	SVM	Decision Trees	KNN
Unigram	86.3	**90**	89.5	90.1
Unigram, Emoticon	86.4	89	89.6	90.1
Unigram, Punctuation	86.6	89.9	89.7	90.1
Unigram, Negation	86.9	89.1	89.7	**90.2**
All-Features	**86.9**	89.9	**90**	90.1

Table 4: F-measure of SVM, Naive Bayes, Decision Tree and KNN using different features

Features	Naive Bayes	SVM	Decision Trees	KNN
Unigram	87.7	90.3	89.6	90.1
Unigram, Emoticon	87.6	89.3	89.7	90.1
Unigram, Punctuation	87.1	**90.4**	89.8	90.1
Unigram, Negation	**87.9**	89.5	89.9	90.2
All-Features	87.3	90.22	**90.1**	**90.2**

Table 5: Precision of SVM, Naive Bayes, Decision Tree and KNN using different features

used this API to collect Tweets for three weeks between December 26, 2013 to January 15, 2014. Data was collected from the Twitter Streaming service, which provides a 1% random sample of all tweets. We have developed a program that filters tweets from online stream of tweets, based on the predefined list of hash-tags.

4.2 Classification Results

We divided the collected data for each class into three equal-sized folds,which used two folds of the labeled data to train a classifier and one fold for testing. Then we learned classifiers out of training data using selected classification algorithms. We used WEKA [25] for Naive Bayes, Decision Tree, and KNN classification, and we used the SVM-light [26] software with a linear kernel to learn SVM classifier.

We measured the accuracy of classifiers based on precision and recall. Also, we calculated the F-measure, which is the weighted harmonic mean of precision and recall.

Tables 4 and 5 present precision and recall of Naive Bayes, Decision Tree, SVM, and KNN using different kind of features, based on 3-fold cross validation.

Table 4 presents the F-measure as a single measure that trades off precision versus recall. As the table shows, the highest accuracy for Decision Trees and Naive Bayes can be achieved by using all the proposed features. However SVM achieved the highest accuracy by using unigrams, while KNN achieved the highest accuracy by using unigrams and

Features	Naive Bayes	SVM	Decision Trees	KNN
Unigram	86.3	**89.7**	89.5	90.1
Unigram, Emoticon	86.4	88.8	89.6	90.1
Unigram, Punctuation	86.6	89.3	89.7	90.1
Unigram, Negation	86.9	88.8	89.6	90.1
All-Features	**87**	89.5	**89.9**	**90.1**

Table 6: Recall of SVM, Naive Bayes, Decision Tree and KNN using different features

negations.

Although Decision Tree classifier provides high accuracy, it is very slow and therefore not practical for big datasets. KNN and SVM run fast and provide the highest accuracy, above 90%.

The accuracy of SVM classification are presented in Figure 6. For the class Unhappy-Active and Happy-Active the highest accuracy can be achieved by using all the features. However for other classes the highest accuracy can be achieved by using unigrams. The classes happy-active and unhappy-active got the highest accuracy. Interestingly, using punctuation features across these two classes increased the accuracy up to 95% and 91% respectively. Across all emotion classes, unigram-trained model gave the highest performance, and among other features punctuations and negations performed second best.

The accuracy of KNN classification based on 3-fold cross validation are presented in Figure 7. As it shows the highest accuracy achieved for the class happy-active. Among all the classifiers KNN achieved the highest accuracy of 90%.

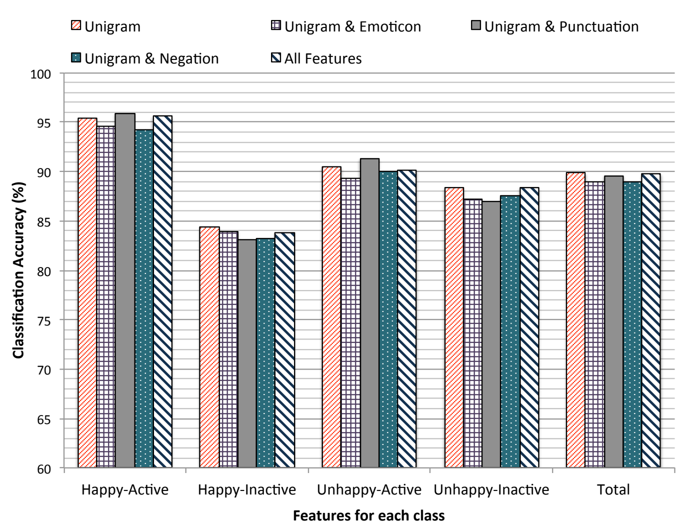

Figure 6: The accuracy of SVM classification using different features

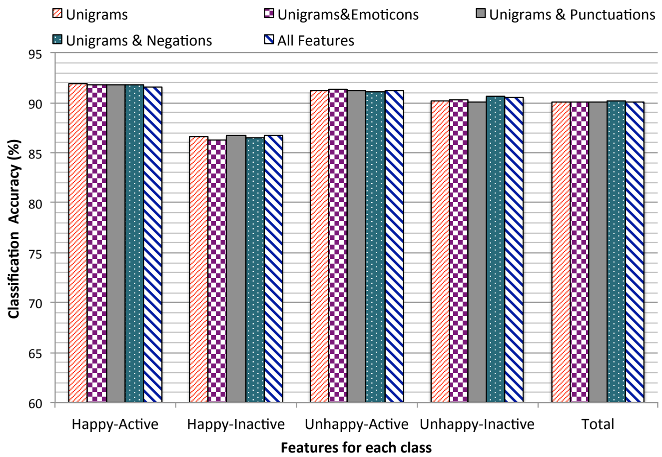

Figure 7: The accuracy of KNN classification using different features

5 Related Work

This section briefly surveys previous works on sentiment and emotion categorization of text.

5.1 Sentiment Analysis in Text

There has been a large amount of research in sentiment analysis, especially in the domain of product reviews and movie reviews. Here we describe those most closely related to our research.

Boia and his colleagues [27] run a study with live users and found that the sentiment of an emoticon strongly coincides with the sentiment of the entire tweet. Thus, emoticons are useful features to predict the sentiment of tweets and should be considered in sentiment classification.

Many researchers applied machine learning methods to detect the sentiment in text. Pang *et al.* [6] classified movie reviews by determining whether a review is positive or negative. They used movies rating indicator, such as the number of stars to automatically label the reviews for supervised learning. They found that standard machine learning techniques definitively outperform human-based classifiers. Among several different types of features they tried including, unigrams, bi-grams, and part-of-speech tags; unigram features turned out to be the most effective.

Some researchers selected twitter as input dataset. Go *et al.* used Western-style emoticons to label and classify Twitter messages according to positive and negative sentiment [7]. They achieved an accuracy of 80% using different learning algorithms including Naive Bayes, Maximum Entropy, and SVM. Other researchers [8, 9] have also used emoticons for labeling their training data.

Another work for sentiment classification on Twitter data has been done by Barbosa and Feng [11]. They exploited existing Twitter sentiment sites for collecting training data, and 1000 manually labeled tweets for tuning and another 1000 manually labeled tweets for testing. They used syntax features of tweets such as re-tweets, hash-tags, links, and exclamation marks.

Kouloumpis and his colleagues [12] investigated using linguistic features for detecting the sentiment of Twitter messages. As labeled data, they used the hash-tag data set, from the Edinburgh Twitter corpus and the emoticon dataset collected by Go *et al* [7]. They achieved an accuracy of 76% by using n-gram, lexicon, and part-of-speech features.

SentiStrength developed by Thelwall *et al* [28], to extract sentiment strength from short informal text. They used a dictionary of sentiment words and their strength measures. They utilized machine learning approach to optimize sentiment term weightings and developed their system using human-classified MySpace comments. SentiStrength could predict positive sentiment with 60% accuracy and negative sentiment with 72% accuracy, both based on a scale of 1 to 5.

5.2 Emotion Analysis in Text

There is much less work on emotion analysis in text compared with sentiment analysis.

Some researchers have explored social media such as Twitter to investigate the potential use of social media to detect depressive disorders. Park *et al* [15] ran some studies to capture the depressive mood of users in Twitter. They studied 69 individuals to understand how their depressive states are reflected in their personal updates. The analysis was conducted in three major steps: (1) surveying the users to identify their depression level, (2) collecting tweets of these users, and (3) comparing the depression levels of users with their language usage in tweets. They found that social media contains useful signals for characterizing the depression in individuals. Their results showed that participants with depression exhibited increase in the usage of words related to negative emotions and anger in their tweets [15].

Another work to diagnose depressive disorders in individuals done by Choudhury *et al* [13]. They measured behavioral attributes including social engagement, emotion, language and linguistic styles, ego network, and mentions of antidepressant medication. Then they leveraged these behavioral features to build a statistical classifier that estimates the risk of depression. Their models showed an accuracy of 70% in predicting depression. They crowdsourced data from Twitter users who have been diagnosed with mental disorders.

Purver *et al* tried to train supervised classifiers for emotion detection in Twitter messages, using automatically labeled data [18]. They used the six basic emotions identified by Ekman [17] including happiness, sadness, anger, fear, surprise and disgust. They used a collection of Twitter messages, all marked with emoticons or hash-tags corresponding to one of six emotion classes, as their labeled data. Their method did better for some emotions (happiness, sadness and anger), than others (fear, surprise and disgust). Their overall accuracies (60%) were much lower than our accuracy.

Another effort for emotion analysis on Twitter data accomplished by Bollen and his colleagues [1]. They tried to find a relationship between overall public mood and social, economic and other major events. They extracted six dimensions of mood (tension, depression, anger, vigor, fatigue, confusion) using an extended version of POMS (Profile of Mood States), a psychometric instrument. They found that social, political, cultural and economic events have a significant, and immediate effect on the various dimensions of public mood.

Some researchers applied lexical approach to identify emotions in text. For example Strapparava and Mihalcea [19] constructed a large lexicon annotated for six basic emotions: anger, disgust, fear, joy, sadness and surprise. In another work, Choudhury *et al* [14] identified a lexicon of more than 200 moods frequent on Twitter. Inspired by the circumflex model, they measured the valence and arousal of each mood through mechanical turk and psychology literature sources.Then, they collected posts which have one of the moods in their mood lexicon in the form of a hash-tag at the end of a post.

Recently, Darmon and his colleagues [29] tried to predict the behavior of users on social media by modeling representations of their previous states as computational processes. They found that most users exhibit only a few latent states of behavioral processing, and any model that is able to capture these states will do well at capturing the behavior of users.

6 Conclusions

We have proposed Emotex, a method of classifying Twitter messages into the distinct emotional classes they express. To define the emotional states of users, we utilized the well-established model of human moods, known as the Circumplex model [10]. We employed Twitter hash-tags to automatically label messages according to emotional classes, and trained classifiers for multi-class emotion detection. Our results suggest that hash-tags and other conventional markers of tweets are useful features for sentiment and emotion classification.

We also compared the accuracy of several machine learning algorithms such as SVM, Naive Bayes, KNN, and Decision Tree for classifying the moods of Twitter messages. We were able to achieve above 90% classification accuracy, while demonstrating robustness across different learning algorithms.

The proposed Emotex approach enables us to classify large amounts of short texts with no manual effort, yet with high accuracy (above 90%). Classifying short texts according to a finer-grained classes of emotions provides rich and informative data about the emotional states of individuals. These data can be used by healthcare professionals for early detection of the psychological disorders such as anxiety or depression.

In the future we intend to analyze the temporal nature of emotions and investigate how they change over time. We are also interested in population level emotion detection in different subgroups like different genders, or ages. In addition, we intend to integrate other pieces of information such as, sleep data, exercise and physical activities, and food information.

References

[1] Johan Bollen, Huina Mao, and Alberto Pepe, "Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena.," in *International AAAI Conference on Weblogs and Social Media (ICWSM'11)*, 2011.

[2] Mike Thelwall, Kevan Buckley, and Georgios Paltoglou, "Sentiment in twitter events," *Journal of the American Society for Information Science and Technology*, vol. 62, no. 2, pp. 406–418, 2011.

[3] Ed Diener and Martin E. P. Seligman, "Beyond money: toward an economy of well-being," in *PSYCHOLOGICAL SCIENCE IN THE PUBLIC INTEREST*. 2004, American Psychological Society.

[4] Ed Diener, *Assessing well-being: The collected works of Ed Diener*, vol. 3, Springer, 2009.

[5] Shigehiro Oishi Ed Diener Ed Diener, "Subjective well-being: The science of happiness and life satisfaction," .

[6] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," in *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing, EMNLP 2002*, Philadelphia, Pennsylvania, 2002, pp. 79–86.

[7] Alec Go, Richa Bhayani, and Lei Huang, "Twitter sentiment classification using distant supervision," *CS224N Project Report, Stanford*, pp. 1–12, 2009.

[8] Alexander Pak and Patrick Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," in *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta, may 2010, European Language Resources Association (ELRA).

[9] Albert Bifet and Eibe Frank, "Sentiment knowledge discovery in twitter streaming data," in *Discovery Science*. Springer, 2010, pp. 1–15.

[10] J. A. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, pp. 1161–1178, 1980.

[11] Luciano Barbosa and Junlan Feng, "Robust sentiment detection on twitter from biased and noisy data," in *Proceedings of the 23rd ACL: Posters*. Association for Computational Linguistics, 2010, pp. 36–44.

[12] Efthymios Kouloumpis, Theresa Wilson, and Johanna Moore, "Twitter sentiment analysis: The good the bad and the omg!," in *International AAAI Conference on Weblogs and Social Media (ICWSM'11)*. 2011, The AAAI Press.

[13] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz, "Predicting depression via

ISBN: 978-1-62561-000-3

social media.," in *International AAAI Conference on Weblogs and Social Media (ICWSM'13)*. 2013, The AAAI Press.

[14] Munmun De Choudhury, Scott Counts, and Michael Gamon, "Not all moods are created equal! exploring human emotional states in social media," in *Sixth International AAAI Conference on Weblogs and Social Media (ICWSM'12)*, 2012.

[15] Minsu Park, Chiyoung Cha, and Meeyoung Cha, "Depressive moods of users portrayed in twitter," in *Proc. of the ACM SIGKDD Workshop on Healthcare Informatics, HI-KDD*, 2012.

[16] James W Pennebaker, Martha E Francis, and Roger J Booth, "Linguistic inquiry and word count: Liwc 2001," *Mahway: Lawrence Erlbaum Associates*, p. 71, 2001.

[17] Paul Ekman, "Basic emotions," *Handbook of cognition and emotion*, vol. 98, pp. 45–60, 1999.

[18] Matthew Purver and Stuart Battersby, "Experimenting with distant supervision for emotion classification," in *Proceedings of the 13th EACL*. Association for Computational Linguistics, 2012, pp. 482–491.

[19] Carlo Strapparava and Rada Mihalcea, "Learning to identify emotions in text," in *Proceedings of the 2008 ACM symposium on Applied computing*. ACM, 2008, pp. 1556–1560.

[20] Jonathan Posner, James A Russell, and Bradley S Peterson, "The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology," *Development and psychopathology*, vol. 17, no. 03, pp. 715–734, 2005.

[21] Xiaolong Wang, Furu Wei, Xiaohua Liu, Ming Zhou, and Ming Zhang, "Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach," in *Proceedings of the 20th ACM international conference on Information and knowledge management*. ACM, 2011, pp. 1031–1040.

[22] Jeffrey T Hancock, Christopher Landrigan, and Courtney Silver, "Expressing emotion in text-based communication," in *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2007, pp. 929–932.

[23] A Gill, R French, Darren Gergle, and Jon Oberlander, "Identifying emotional characteristics from short blog texts," in *Proc. 30th Ann. Conf. Cognitive Science Soc., BC Love, K. McRae, and VM Sloutsky, eds*, 2008, pp. 2237–2242.

[24] Thorsten Joachims, "Text categorization with support vector machines: Learning with many relevant features," in *Proceedings of ECML-98, 10th European Conference on Machine Learning*. 1998, pp. 137–142, Springer Verlag, Heidelberg, DE.

[25] Eibe Frank, Mark Hall, Geoffrey Holmes, Richard Kirkby, Bernhard Pfahringer, Ian H. Witten, and Len Trigg, "Weka-A Machine Learning Workbench for Data Mining," in *Data Mining and Knowledge Discovery Handbook*, Oded Maimon and Lior Rokach, Eds., chapter 66, pp. 1269–1277. Springer US, Boston, MA, 2010.

[26] Thorsten Joachims, "Making Large-Scale SVM Learning Practical," in *Advances in Kernel Methods - Support Vector Learning*, Bernhard Schölkopf, Christopher J.C. Burges, and A. Smola, Eds., Cambridge, MA, USA, 1999, MIT Press.

[27] Marina Boia, Boi Faltings, Claudiu Cristian Musat, and Pearl Pu, "A :) is worth a thousand words: How people attach sentiment to emoticons and words in tweets.," in *SocialCom*. 2013, pp. 345–350, IEEE.

[28] Mike Thelwall, Kevan Buckley, Georgios Paltoglou, Di Cai, and Arvid Kappas, "Sentiment strength detection in short informal text," *Journal of the American Society for Information Science and Technology*, vol. 61, no. 12, pp. 2544–2558, 2010.

[29] David Darmon, Jared Sylvester, Michelle Girvan, and William Rand, "Predictability of user behavior in social media: Bottom-up v. top-down modeling.," in *SocialCom*. 2013, pp. 102–107, IEEE.