Framework For Emotion Detection And Classification of English and Non-English Tweets

Anup Prasad
Indiana University Bloomington
School Of Informatics And Computing
anprasad@umail.iu.edu

Mayank Jaglan
Indiana University Bloomington
School Of Informatics And Computing
mjaglan@umail.iu.edu

Abstract

Social media platforms, like Twitter, have quickly become popular as a means of sharing messages, primarily in the form of text. These text can be news, feelings, thoughts, observations, life experiences and moments. Any data shared exhibits either positive vibes, negative vibes or it can be neutral content [3]. While positive vibes and negative vibes are formed from the opinions attached to the content, neutral contents are mere facts explained with no sentiments. In this project, we have worked on text only data obtained from twitter platform to discover, analyze and verify 8 different kinds of basic emotions as identified by Pultchik. The text data covers English as well as Non-English languages. We have built a framework that can mine the users emotions on live as well as past twitter feeds. For non-English tweets, we first convert them into English using Bing translator API. Then with the help of an emotion tagger, we convert each tweet into a vector of eight basic emotions. These vector representation of tweets are then fed into K-means, Naive Bayes and SVM to produce label corresponding to one of the basic emotion family. Towards the end, we have done a comparative study of the performance of different classifiers.

1 Credits

Bing Package,
github.com/openlabs/Microsoft-Translator-Python-API

Twitter Package,
github.com/geduldig/TwitterAPI

Google Geo Package,
pypi.python.org/pypi/geocoder

Scikit-learn Package,
github.com/scikit-learn/scikit-learn

Classroom readings and discussions (Fall 2015, Social Media Mining, ILS-Z639 by Professor Muhammad M. Abdul-Mageed (http://ella.slis.indiana.edu/ mabdulma/)

2 Introduction

Social medial platforms like Twitter and Facebook are rich source of information. This information can be about events or news or any other moments shared online with everyone. Often people express their opinion on these social media platform about these events, which is rich in emotion. In our work we are identifying emotions present in tweets related to a particular major event and trend in a particular city, so that we can know about sentiments of population related to some event e.g. what was the sentiment of masses when Hillary Clinton or Donald Trump is in news. This emotion mining has numerous practical applications for example a restaurant brand like McDonald can visualize how customers feel about their service or food or pricing.

3 Related Work

While Emotion classification has been an area of research for past three decades, emotion mining is relatively new and can be attributed to the rise of social media platforms which has added a new dimension to the freedom of expression.

It is vital, for emotion mining, to identify various categories and families of emotions. Emotions have been divided into six basic families by Paul Ekman(1992)[3].According to him, these basic emotion families are discrete and are basic in the sense that other complex emotions are a combination of two or more basic emotions. The six basic emotion families identified by Ekman are Anger, Sadness, Disgust, Fear, Enjoyment, Surprise .One characteristic of Ekman's classification is that these emotion families are skewed towards negative emotions .i.e majority of them represents a negative emotion.

Another classification was proposed by Plutchik (1980)[4], he has created eight emotion families where each positive emotion is having an opposite negative emotion. Thus, in the classification by Plutchik, there is a pair of positive and negative emotions. These pairs are Trust and Disgust, Joy and Sadness, Surprise and Anticipation, Fear and Anger. He has also created a wheel of eight emotions where emotions are positioned according to their valance, intensity level and composition (i.e two basic emotions can be mixed to produce a third emotion).Emotion pairs are vital for the task of emotion mining on text base medium because emotion words can appear in association with model verbs and adverbs like Not, But etc. which shifts valance of an emotion.These emotion pair helps in achieving the task of valance shift .

Once a classification is selected, the next task is to develop a model to perform emotion mining on the text. There are different models that have been proposed to find emotions in a review .Lionel Martin and Pearl Pu have used GALC[5], a general lexicon of emotional words associated with a model representing 20 different categories, to extract emotions from the review's text and applied supervised classification method to derive the emotion-based helpful review prediction. They have used Support Vector Machines, Random Forests and Naive Bayes for this classification.In their work they have achieved important improvement using SMOTE algorithm on Random Forests for the majority of the feature sets on all their datasets compared with undersampling and a smaller significant improvement on Support Vector Machine.

There are various challenges in sentiment analysis as identified by Bing Liu[6] .The first challenge is to identify if a review is subjective or objective. In our project ,it was difficult to find if a review had factual information or if it contained some sentiment or emotions in it. While it was easy to work on subjective reviews which contained explicit emotions or sentiment, working on objective reviews which contained implicit emotions or no emotion at all was a challenging task and required domain knowledge based machine learning. While working on the project, we found that unsupervised learning techniques could be easily used to identify emotions in subjective reviews with the help of a POS tagger. As mentioned by Bing Liu in the publication ,"Sentiment Analysis and Subjectivity" various bi-grams containing adjectives and adverbs are good indicators of subjectivity and opinions that can be extracted with the help of a POS tagger. Another major challenge was to identify the orientation of sentiment. It was found that the sentiment of words changed their orientation when they were used in association with some other word .e.g 'not happy' or 'rarely sad'.

The presence of reviews or opinion in various dialects and languages poses another set of challenge to emotion mining on social media platform. As pointed out by Anna Katrine Jrgensen, Dirk Hovy, and Anders Sgaard in their paper[7], that dialect features- which otherwise doesn't creep into formal writing, is prevalent on social media .This is the reason that POS taggers which are trained on formal writing, performs miserably when these dialect features creeps into sentences.

In order to build an automatic emotion classifier, as proposed by Saima Aman and Stan Szpakowicz[2], sentences can be converted into a vector of emotions .This emotion vector contains either frequency of occurrences of various emotion words representing different emotion families or they contain weighted values of their occurrence.One way to identify emotion words in a sentence is to create a Bag of emotion words for each emotion family. These emotion words can be taken from two publicly available lexical resources the General Inquirer and WordNet-Affect . The next step in supervised learning is to gather training data set, which can be labeled automatically or can be labeled manually. In our project, we are generating automatically labeled training data by crawling over live tweeter feeds.

The concept of valance shift as addressed by Livia Polanyi and Annie Zaenen in their paper[1] plays an important role while performing sentiment analysis.There are a number of ways in which the basic valence of individual lexical items may be strengthened or weakened or changed by context provided by the presence of other lexical items, the genre type and discourse structure of the text and cultural factors. While building a function which converts a text into a vector of emotions, contextual valance shift must be considered. In our project, we have addressed these phenomenon to a larger extent.

Next challenge is how to gather proper training data. As suggested by Barbara Plank and Dirk Hovy[8] (in How to Get 1,500 Personality Tests in a Week), we generate large number of self identified training tweets in a week. We then manually check all files and remove all tweets that contain more than one emotion category. This typically relates to mixed or more complex emotion, In the end, our collection contains 12000 distinct tweets with emotion vector and emotion label for training purposes.

4 Task of Classifying Tweets

Our task was to detect 8 basic emotions, as proposed by Plutchik, in the live twitter stream and classify tweets as 'anticipation', 'enjoyment', 'sad', 'disgust', 'anger', 'surprise', 'fear', 'trust' or No Emotion. This classification is done using supervised machine learning : - Naive Bayes and SVM and unsupervised machine learning :- K-means.

We have created 8 bags of words, each of which represents one emotion family .The bag of words are mutually exclusive i.e word from one bag cannot be present in another bag.

$$\{Enjoyment\} \cap \{Fear\} \cap \{Sad\} \cap \{Disgust\} \cap \{Anger\} \cap \{Surprise\} \cap \{Fear\} \cap \{Trust\} = \emptyset$$

Text Emotion Classification (TEC) System which we have built can work on multiple languages. Every non-English tweet is first converted into English sentences with the help of Bing translator (for now), then an emotion tagger converts each tweet into a vector of 8 basic emotions. The approach of using a translator is based on the argument that in any text we just need English words that safely fall into one of the 8 basic emotion category related bag of words (BOW). These vectors are then fed into SVM and Naive Bayes, which gives us emotion label for the tweet. K-means approach work differently from SVM and Naive Byes , It needs entire data set to work on. The output of K-means is K clusters, these K clusters are then assigned a label based on the frequency distribution of emotions in the cluster.

A comparative study of the efficiency of these three classification methods has been performed. We found that efficiency of SVM's One vs Reset Classifier is the highest followed by slight drop for Naive Bayes Classifier. While K-means's efficiency has been modest. Source code for this study is present at https://github.iu.edu/mjaglan/TextSentiment.V1.a

5 Data set

5.1 Data Query

To generate training data set, we used 100 words from each bag.These words were first converted into hashTags and then fed into our tweet retrieval system (TRS).The TRS keeps on listening to the live tweet stream and captures any tweet with the hasTag which is passed to it. Labeling of training data set is automated, when a tweet is retrieved it is assigned a label which is one of the basic emotion family to which hashtag word belongs.

Tweets	#Tag	Label
Tweet1	#FURY	Anger
Tweet2	#Expect	Anticipation
Tweet3	#Odium	Disgust

Table 1: Sample Tweets with label

Our TRS framework can be used to retrieve tweets based on Geolocation, specific language, hashTags and any key Word.

5.2 DATA CLEAN

The tweets which our TRS system retrieves contains lots of garbage value.These values or connectors have to be removed from training data set before it is passed to emotion tagger to get mapped to the vector representing basic emotion. URL, new-lines, tabs, hash characters,êtc.

Garbage value	Example	re-Expression
URL	https://t.co/eZfgqKv0yR	https?:.[$\backslash r \backslash n$]*
$\backslash n$	so much anger and hurt$\backslash n$	[$\backslash n$]
$\backslash t$	Personally $\backslash t$ I think it's..	[$\backslash t$]

Table 2: Sample Tweets with label

5.3 Corpus Collection

Total of 6000+ tweets were collected for training purpose.These tweets are labeled and are associated with vector corresponding to basic emotion family. Other meta data associated with the training set are as follows.

ID	TIME	LAN	ANNOT	VECT
6701	11/27/2015 08:16	en	anger	1 0 0 0 0 0 0 0

Table 3: Sample Tweets with label

6 Methodology

At the core of our Text Emotion Classification (TEC) is emotion tagger. Emotion tagger takes a sentence and converts it into a vector e.g.

Text	Vector
im angry again ugh	1 0 0 0 0 0 0 0
A cowardly leader is the most dangerous	0 0 0 0 1 0 0 0
Victory of your pen. Gives me immense pleasure to imagine their frightened faces. They are bloody SCARED!!!	0 0 0 1 2 0 0 0

Table 4: Sample Tweets with label

These vectors are then used to train SVM and Naive Byes. In our implementation, we have used linear kernel for SVM.Once SVM and Naive Bayes are trained then live tweets can be fed into them to determine content of its emotion type.

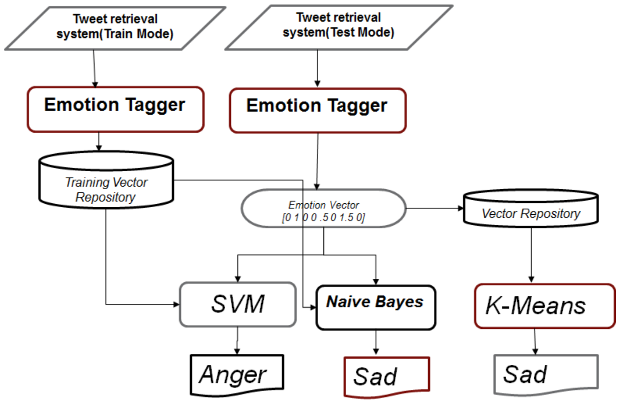

Image 1 : Data Flow

6.1 Design and Working Of Emotion Tagger

The emotion tagger (ET) which converts each tweet into a vector of emotion works at the syntactic label of a sentence.As a result it can detect emotions in subjective tweets or reviews because emotions are explicitly mentioned in them.However it fails to recognize implicit emotions which may or may not be present in objective reviews or tweets. As mentioned earlier, we have created 8 mutually exclusive bags of words. These words are most frequently used to describe or express emotion. They are hand picked EMO words from "General Inquirer".

The ET can detect contextual valance shift which takes place under certain conditions, These are described below below .

1. It can detect valance shift due to presence of negation like NOT.
2. It can detect intensifiers like Very , So ,Deeply .
3. It can detect presupposition items like barely ,even .
4. It can detect words like rarely , seldom which decrease intensity of emotion .
5. Certain words like "neither nor" which nullifies intensity of emotion eg. I am neither sad nor happy ,can be detected by the tagger.
6. It can detect uni-gram, bi-gram and multi-grams.

ET handles all these situations differently and assigns intensity to each emotion according to following rules .

Context	Example	Intensity
Negation	NOT	(-)ve
Intensifiers	Very	1.5
Presupposition	barely	0.5
Nullifiers	neither nor	0

Table 5

Whenever ET encounters a negation which is associated with emotion word it changes the emotion family to it's opposite emotion family e.g. "I am not happy(Enjoyment family)" is treated as "I am sad(Sad family)".

In order to detect bi-grams and multi-grams ET tags each sentence with the help of NLTK POS tagger. These patterns can be present in following forms.

Word Extent	Example	POS pattern
Bi-Gram	rarely happy	RB JJ
Multi-Gram	inability to perform good	NN TO VB JJ

Table 6

The over all working of ET is described in following diagram .

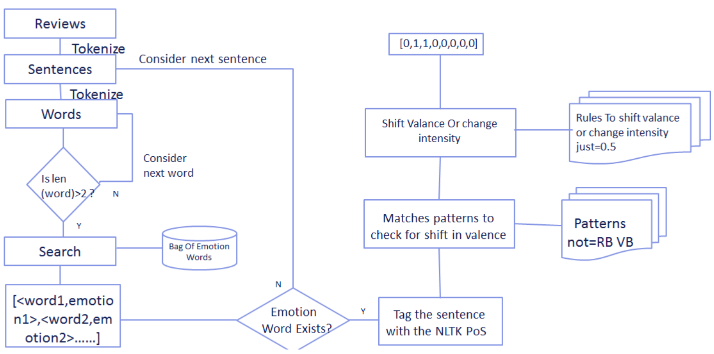

Image 2: Emotion Parser

6.2 K-Means For Classification

K-Means is a clustering algorithm. But in our project, we have used it for classification.It is unsupervised learning algorithm so it doesn't require any training data set. We run K-Means algorithm on the data set collected for training, the output is K clusters. To assign a label to these clusters, we calculate the frequency distribution of emotions of each cluster. The emotion family whose frequency is highest is then assigned as the label to the cluster.So label of a tweet is same as the label of the cluster to which it belongs. The entire labeling process is repeated n-Times, in our case, it is 17. This is to mitigate the effect of random initialization done in K-Means.

The entire process is explained in following diagram .

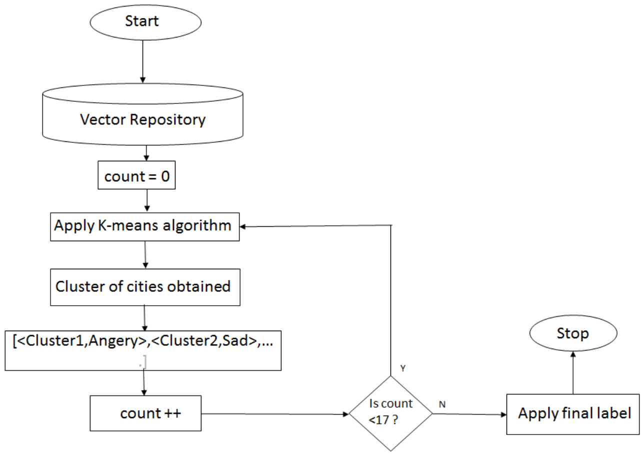

Image 3: Emotion Parser

7 Result

Accuracy Baseline 1: 23.45 percent
Accuracy Baseline 2: 39.6 percent

Algorithm	Accuracy	Precision	Recall
KMeans(K=8)	40.1	41.31	40.40
KMeans(K=8+1)	40.00	41.00	40.0
NBC(5-F)	64	58	64
NBC(10-F)	64	58	54
SVC(5-F))	66	56	66
SVC(10-F)	65	56	56

Table 7:Performance Table

Accuracy Baseline 1 mentioned above has been calculated from training data set where Anger was the most prevalent emotion. For Training, we have used 13000+ tweets.
Anger: $3220/13731 = 23.45\%$

Accuracy Baseline 2 above has been referred from the work of Soumaya Chaffar and Diana Inkpen[9].

It is observed that SVC approach (with 5-Fold validation) is the best performer. Following is the classification report:

Emotion	precision	recall	f1-score	support
anger	0.80	0.91	0.85	254
anticipation	0.00	0.00	0.00	79
disgust	0.63	0.84	0.72	73
enjoyment	0.60	0.86	0.70	171
fear	0.61	0.79	0.69	173
sad	0.66	0.61	0.64	199
surprise	0.56	0.51	0.53	83
trust	0.00	0.00	0.00	87

Table 8:Classification report of Linear SVC(5-Fold)

It is evident from above table that NBC and SVM works better than K-Means. Also 5-fold cross validation performed better than 10-Fold. And K-means performance doesn't change with K=9 and K=10.

One of the practical application of the emotion detection framework, which we have built, is to know the general mood of public about something. Here we have tried to find what people feels about the USA presidential candidates of 2016.

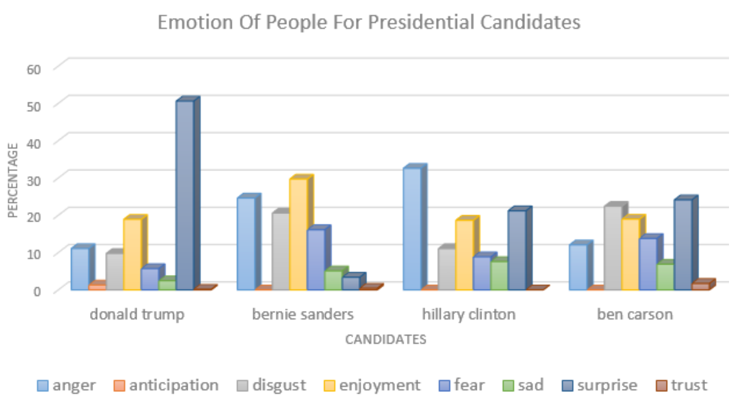

Image 4 : Emotion Of Public

We have collected around 400 tweets for each of the four candidates. We remove the tweets that contained no emotion or more than 5 unique emotions. And it is interesting to note that people do not trust any of them !

8 Appendix

1 Getting good quality of data remains the most important part.We are often getting tweets having only URLS , so data pre-processing still remains one of the most crucial steps which we need to improve.

2 We are only exploring explicit emotion present in a tweet. In our future work, we need to devise a method to detect implicit emotions present in objective sentences.

3 In present implementation we are treating each sentence as an individual unit. In our future work, we will include the association between sentences as well.

4 Emotion tagger, which is the core of our emotion detection framework, works at the syntactic level only.In future, we propose to add semantic level parsing as well.

5 Bag of words needs to be more exhaustive. In future we will add more words, describing emotions, to these bags.

9 Attendance in Class

Mayank Jaglan : No class missed
Anup Prasad : 2 classes missed
Reason : Was having midterms.

Acknowledgments

We would like to thank Dr.Muhammad Abdul Mageed, for his undaunted guidance and support, without which we would not have been able to finish this project on time.Reviews provided by him has always inspired us to set our goal higher.Also extra miles he has taken to guide and support us has always galvanized us to work harder.

References

[1] Livia Polanyi and Annie Zaenen (2006). Contextual Valence Shifters. *Computing attitude and affect in text: Theory and applications, 1-10.*

[2] Saima Aman and Stan Szpakowicz (2007). Identifying Expressions of Emotion in Text. *In Vclav Matousek and Pavel Mautner, editors, Text, Speech and Dialogue, volume 4629 of Lecture Notes in Computer Science, pages 196-205. Springer Berlin / Heidelberg.*

[3] Ekman, P. (1992). An Argument For Basic Emotions. *Cognition and emotion, 6(3-4), 169-200.*

[4] Mohammad, S. M., Turney, P. D. (2013). Crowdsourcing a Word-Emotion Association Lexicon, *Computational Intelligence, 29(3), 436-465.*

[5] Lionel Martin and Pearl Pu (2014). Prediction of Helpful Reviews Using Emotions Extraction. *AAAI Conference on Artificial Intelligence Twenty-Eighth AAAI Conference on Artificial Intelligence*

[6] Bing Liu. (2009) Sentiment Analysis and Subjectivity. *Handbook of Natural Language Processing, Second Edition,(editors: N. Indurkhya and FJ Damerau).*

[7] Anna Katrine Jrgensen, Dirk Hovy, and Anders Sgaard (2015). Challenges of studying and processing dialects in social media. *In Proceedings of the Workshop on Noisy Usergenerated Text (pp. 9-18).*

[8] Plank B., Hovy D. (2015). Personality Traits on Twitter–or–How to Get 1,500 Personality Tests in a Week. *In 6th Workshop on Computational Approaches to Subjectivity, Sentiment, and Social Media Analysis (WASSA 2015) (p. 92).*

[9] Soumaya Chaffar and Diana Inkpen (2011). Using a Heterogeneous Dataset for Emotion Analysis in Text. *Proceeding Canadian AI'11 Proceedings of the 24th Canadian conference on Advances in artificial intelligence Pages 62-67*