CIS 630 Project 2 Report: Twitter Affect

Tao Feng, Yayang Tian, Chun Chen

May 7, 2013

1 Introduction

We present a practical approach to classify tweets into five emotions classes. By incorporate emotion dictionaries with our natual language processing model, no manual intervention is needed to summary what people feel on Twitter.

1.1 Contribution

1. **Tweets Corpus:** Presented a method for automatically collecting recent tweets with different emotions. Created a large tweets corpus consisting of five emotions: "happy, sad, angry, afraid, ashamed".

2. **Six-class Classification Model:** Adopted various methods for tweets affect classification and outperform baseline approach by 21.197%. Conducted experimental evaluations on real-time Tweets and showed the importance for stemming, affect dictionary, smiley, information gain, and SVM in five-class classification.

3. **Web Application:** Build a web application that can classify and summarize emotions on Twitter in real-time based on user-specified keyword.

1.2 Formulation and Challenges

1. **How to collect our data?** There are many existing tweets corpus. For example, E Most of them are confined to positive, neutral, and negative emotions. We need to find our ways that best suits five-class classification.

2. **How to label our data?** Manual labeling from experts might be a ideal way. But given large corpus we need to collect, it is not practical. We need find ways to correctly label tweets before training.

3. **Which features for Tweets?** Tweets are different from traditional text in that it is not formal and restricted to only 140 words. This on one hand brings about noise to training, on another exposes special features that may be informative for emotion discrimination.

1.3 Related Work

1. Broad overview for sentiment analysis (Pang & Lee, 2008) [4]
2. Microblog with SVM and CRF (Yang et al., 2007)
3. Emoticons like ":-)" (Read, 2005)[12]
4. Tweets with Distance Supervision (Go et al., 2009)[5]
5. Six-way emotion classification (Ansari, 2010)
6. Multi-class emotion detection Distance Vector Supervision (Matthew, 2010)

2 Data Collection

Up to now, there is no existing Tweets corpus that are categorized into five emotion categories: Sad, Angry, Happy, Ashamed and Afraid. Therefore we decided to collect our own data.

2.1 Dictionary-based Gold Standard

Given the words-emotion matrix from the Emotion Intelligence, We collected a tweets corpus and divided them into five emotion classes based on lexicon occurrence in Emotion Intelligence Dictionary.

Affect	Words
Happy	Elated, Excited, Overjoyed, Thrilled, Exuberant, Ecstatic, Passionate, Cheerful, Gratified, Relieved, Satisfied, Glad, Contented, Pleasant, Pleased, Mellow
Sad	Depressed, Agonized, Dejected, Hopeless, Sorrowful, Miserable, Heartbroken, Somber, Distressed, Melancholy, Unhappy, Moody, Upset, Disappointed, Dissatisfied
Angry	Furious, Enraged, Outraged, Irate, Seething, Loathsome, Betrayed, Upset, Frustrated, Agitated, Disgusted, Annoyed, Uptight, Resistant, Irritated, Touchy
Afraid	Terrified, Horrified, Petrified, Fearful, Panicky, Frantic, Apprehensive, Frightened, Threatened, insecure, Uneasy, Intimidated, Cautious, Nervous, Worried, Timid, Anxious

2.2 Twitter Streaming API

We developed a Tweets crawler using Twitter streaming API[19]. It downloaded tweets that contain any words from the emotion chart and labeled them as corresponding emotions. We crawled 42416 tweets and divided them into 32934 training data and 9492 testing data.

Affect	Training data	Testing data
Happy	9168	1587
Sad	7639	2529
Angry	6138	1728
Afraid	5447	1818
Ashamed	4532	1830
Total	32924	9492

2.3 Preprocessing

Before expanding the corpus, we preprocessed the tweets. This is important especially to informal microblog.

Method	Accuracy improved
Lowercase	Great
URLs	Not Much
@Username	Not much
Puctuations	Some
Stemming	Great
Reduce "happppppy" to "happy"	Not much

Here are two examples for preprocessing.
(1) Before: @ Msdebramaye I heard about that contest! Congrats girl! http://flickr.com/td3zfa
After: someUSER i heard about that contest! congrats girl!! someURL
(2) Before: Good night #Twitter and #TheLegionoftheFallen. 5:45am cimes awfully early!
After: good night twitter and thelegionofthefallen. 5:45am cimes awfully early!

2.4 Correctness of the labeling

Before training, it's helpful to justify the correctness of automatic labeling, we trained five binary classifiers and derived the following results:

Emotion	Example words
Happy Not Happy	85.3666%
Angry	83.2807%
Sad	75.8639%
Ashamed	82.1745%
Afraid	84.5133%

The above table shows that each emotion can be well distinguished from others. As can be seen below these five binary classifier could help create an emotion profile[4].

3 Baseline

The baseline for our experiment unigrams with Naive Bayes classifier. We know that Naive Bayes is widely considered as a simple but powerful classifier, so our baseline was relatively stronger, compared to the common baseline using emotion words counts. Plus, in contrast to (Alec, 2010) and (Alexander Pak, 2010) who concluded that their best model for two-class or three-class tweets classification was given by Naive Bayes [5, 14], we found SVM gave us better results for emotion recognition. We utilized scikit learn[15] with Gaussian Kernel and got accuracy of **28.9506%**.

4 Experiment

We then switched the machine learning methods to SVM(Vapnik, 1995), with the help of libsvm developed by (Chang and Lin 2010). We explored useful features for tweets and reduce noisy ones[16]. We do not do any feature selection here and the accuracy we got is **46.46%**. Among the top three classifier[10] Naive Bayes, Max Entropy, and SVM, the last one seemed to gave us best performance.

4.1 Stop List

The first step for refining features was to excluded words that had low inverted document frequency. Most of these stop words were blog-related: hashtag, URL and retweet.

someTAG	someURL	someUSER	#	[	]	&	lt	&lt

4.2 Part of Speech

Given that including the part of speech features actually gives a drop in performance for tweets(Kouloumpis 2010), and (Alec 2010), we did not do POS tagging here. The reason of bad performance for POS may due to the accuracy of the POS tagger on the tweets or POS tags are less useful on microblogging data[20, 5].

4.3 PMI

We then computed the point wise mutual information[1] of the tweets for each emotion category, similar to(Peter, 2012) approach, as a way to further reduce irrelevant features. The words that had high PMI are as follows:

Emotion	**Top PMI**	**Words**
Happy	1.2793	remastered, while, tipsy, trotar, all-county, karley, nadie, figs, wthru, portugaaaaal, pagmaytime
Sad	1.4606	tatum, sentir, push-ups, lakers, vacaciones, ring, wahh, c, immune, sf, directioner, tastes, rooting
Ashamed	1.9827	deli, win-, jedward, disturbed, delicatessen, agram, passenger, rents, ahorita
Angry	1.6794	mecahin, vshare, aggression, underdogs, nooooooooooooo, uofa, shenanigans, count, data, disturb
Afraid	1.7988	soonerrr, shura, tommarow, omigosh, yu, semakin, sanderson, charted, thunder

These words were not quite indicative so we did not take these features into account.

4.4 TF-IDF weights

Given the promising result of the SVM, we then tried to do some feature selection here. The first method we tried was the TF-IDF. For each emotion category, we computed the TF-IDF for each emotion and got the top words for each emotion. The sample result is as follows:

Emotion	**Emotions words**
Happy	Holi, followers, usa, *-*, feliz, host, mexico, prom, yay
Sad	bulls, streak, lakers, duck, dynasty, heats, struggling, seed
Ashamed	fools, silly, fooled, guilty, ure, hot., embarrassing, Chelsea, prank, fool
Angry	angry, irritated, pissing, annoys, pisses, cares, wtf., punching, annoyed, fix, plans
Afraid	tryouts, scared, mri, audition, spider, dory, messi,scares, 2015, results

Again, these features was not indicative of emotions, so we didn't use them for the task.

4.5 Icon Symbol

We then explored useful features that are special to tweets. There are some Unicodes that used in tweets very frequently nowadays. After extracting them and taking them as features, we got accuracy of **46.40%** which was slightly worse than the pure unigram method, indicating that these Unicode symbols was somewhat noisy in tweets emotion identification task.

4.6 Emoticon

Emoticon such as :) and : (are becoming increasingly common on Twitter. We added a feature indicating how many emoticons appeared in the tweet, which gave us an improvement of accuracy to **46.74%**. This demonstrated that emoticons are strong indicator about how people feel.

4.7 Lancaster Stemmer

After applying the Lancaster stemmer from the nltk[17] tool kits, we gain improvement for the accuracy to **49.06%.** This demonstrated that normalization did help precision/recall.

4.8 Words Only

Another method we tried was to simply use word and ignore everything else including emoticons, Unicode symbols, punctuations etc. However, this decreased the accuracy to 43.08%, which indicated that the extra features that were not pure English words, conveys a lot of information that couold help the classifier to identify emotions.

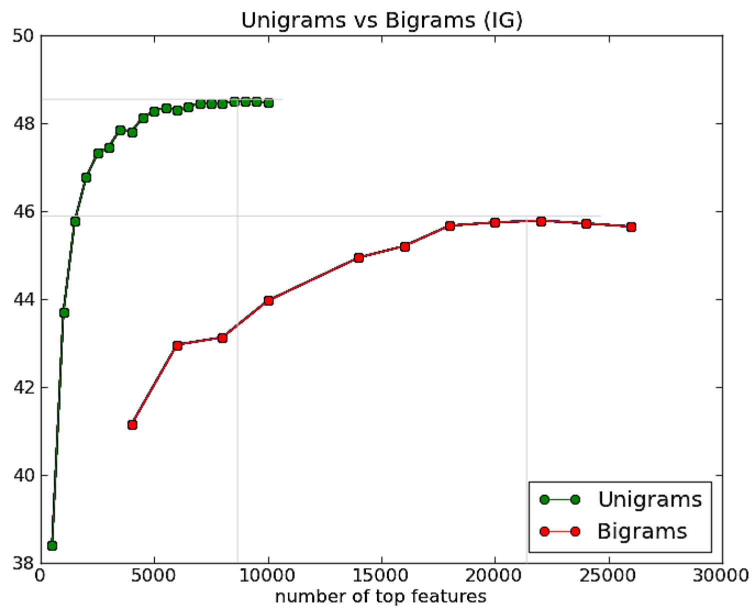
4.9 Stemmer and Emoticons

Here we combined the two methods that give us an improvement. We stemmed the words and added emoticon features at the end. We got accuracy of 49.32%, which beated either of them alone. This showed that stemmer and emoticons were both good ways to extract useful features for tweets.

4.10 Information Gain

In order to get an idea of how IG affected the accuracy like (Barbosa, 2010) did, we tried various number of IG features both Unigram and Bigram[13], the performance comparison are shown below.

Features	Unigram	Bigram
top 2000	46.7657%	37.9372%
top 4000	47.8087%	41.1504%
top 6000	48.2933%	42.9625%
top 8000	48.4408%	43.1311%
top 10000	48.4724%	43.9633%



We can see from the above graph that the increase with the number of features can actually increase the accuracy. This to some extent indicates that information gain servers as an effective feature selection methods here.

If we combined unigrams and bigrams. Mixing 4500 Unigrams and 8000 Bigrams gave us accuracy of 49.4732% while mixing 8000 Unigrams and 16000 Bigrams resulted in accuracy of 49.9157%.

Model	Best Accuracy
Unigrams	48.4724%
Bigrams	45.649%
Unigrams & Bigrams	49.9157%

So we concluded that the best performance were given by combing unigrams and bigrams. On the other hand, we could combine them at the decision level, which would leave for future work.

4.11 Emotion Dictionary

By making use of words' prior polary[9], we calculated how many words appear in each of the directory and train the classifier, we also combined four of the directory features, the result is as the following table:

Dictionary	Num Ratio	Accuracy
Birdy's	3057/9492	32.2061 %
Bing Liu's	4566/9492	48.1037 %
FrameNet	4627/9492	48.7463 %
MPQA	4682/9492	49.3257 %
WordNet	4693/9492	49.4416 %
Mix All	4687/9492	49.3784 %

We can see from the results that Bing Liu's list[6] is slightly worse than others while WordNet[18] gave us best performance for dictionary-based approaches.

4.12 Emotion Profile

Finally added neutral class to our data to examine if there were plenty of tweets with simple neutral emotions. We adopted emotion profile(Emily, et al. 2010) which was basically five binary classifier trained using SVM. If all of the classifier stated that the tweet didn't belong to them(say, with probability lower than 0.6). Then we considered it as neutral. The final prediction of emotion was the one with max confidence or hyperplane distance[7].

Binary	Accuracy
Happy or Not	85.3666%
Angry or Not	83.2807%
Sad or Not	75.8639%
Ashamed or Not	82.1745%
Afraid or Not	84.5133%

Among the total 9492 tweets, 4509.0 are predicting correctly. The final accuracy of emotion profile is **47.5032%**. It performed well, yet slightly less than the previous model.

5 Best Result

Up to now, we see three feature selection methods that can actually improve the accuracy, which is Stemmer, emoticons and WordNet dictionary. This result accorded with the experimental results by (Efthymios and Wilson, 2011)[20].In order to fully diverse the feature space, we combine top 8000 Unigram words, and 16000 bigram phrase, with highest IG. Emoticons and Word word count are also included. Finally we got the best result: 50.1475% (4760/9492). The best model thus could be summarized as follows:

Step	**Methods**
Data Collection	Language filter Content seen
Preprocessing	Replace retweets Stemming ToSmallCase
Features	Unigrams & Bigrams WordNet Dictionary Emoticon
Machine Learning	SVM
Cross Validation	**50.1475%**

5.1 Confusion Matrix

To visualize the performance of our best model, we derived the following confusion matrix:

x axis: prediction	Happy	Angry	Sad	Ashamed	Afraid
Happy	77.9%	29.6%	7.6%	5.0%	6.5%
Angry	11.8%	48.0%	23.6%	5.3%	8.5%
Sad	21.2%	16.3%	45.5%	9.5%	7.6%
Ashamed	22.6%	12.5%	18.4%	34.9%	7.8%
Afraid	21.7%	9.9%	14.4%	5.6%	48.5%

From the table, we can see:
(1) "Happy" had a percentage of 29.6% to wrongly classified as "Angry". This may because we did not take negation into consideration. Negation such as "Not" are used very frequently in tweets.
(2) The two negative emotions "Angry" and "Sad" were easy to get confused. This further demonstrated that it is challenging to distinguish subtle emotions for same polarity.
(3) "Ashamed" and "Afraid" also had high probability to be considered as happy. This may because there existed plenty of negation in tweets which we hadn't considered.

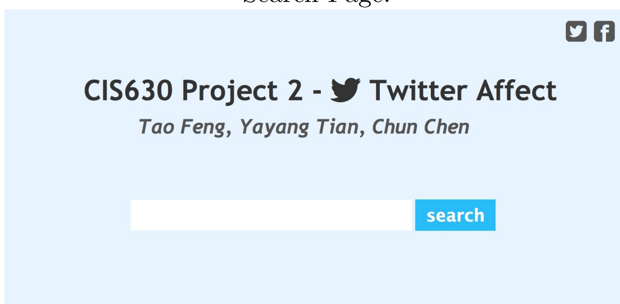
5.2 Manual Annotation

To illustrate our model, we also manually annotate 200 random some tweets. The resulting accuracy for average annotation from three CS graduate students were **42.500%**. Compared to 50.14750% of our model. It showed that the accuracy for human was possibly even lower than our classification system. So in the future we need to find ways to manually label our tweets corpus instead of using dictionary as guideline.

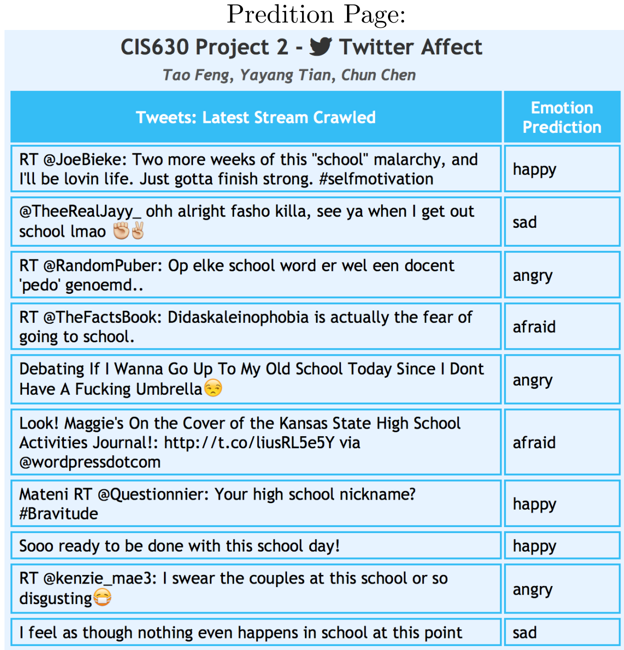
6 Web Application

To better illustrate our work, we also build a web application in Python/Django. It searches for tweets around you based on your query and displays tweets along with their predicted emotions. Similar work includes Twittr and Twendz, Twitter sentiment and TweetFeel, all of which are limited to polarity discrimination. Our application is a pilot one that is able to classify tweets into five emotions.

Search Page:



Predition Page:


CIS630 Project 2 - Twitter Affect

Tao Feng, Yayang Tian, Chun Chen

Tweets: Latest Stream Crawled	Emotion Prediction
RT @JoeBieke: Two more weeks of this "school" malarchy, and I'll be lovin life. Just gotta finish strong. #selfmotivation	happy
@TheeRealJayy_ ohh alright fasho killa, see ya when I get out school lmao	sad
RT @RandomPuber: Op elke school word er wel een docent 'pedo' genoemd..	angry
RT @TheFactsBook: Didaskaleinophobia is actually the fear of going to school.	afraid
Debating If I Wanna Go Up To My Old School Today Since I Dont Have A Fucking Umbrella	angry
Look! Maggie's On the Cover of the Kansas State High School Activities Journal!: http://t.co/liusRL5e5Y via @wordpressdotcom	afraid
Mateni RT @Questionnier: Your high school nickname? #Bravitude	happy
Sooo ready to be done with this school day!	happy
RT @kenzie_mae3: I swear the couples at this school or so disgusting	angry
I feel as though nothing even happens in school at this point	sad

7 Conclusion

We have presented a system for classifying real-time tweets into one of the five emotion categories, which is a relatively new area of research. We found that blog-specified features such as hashtag and emoticon[8] could provide evidence for affect classification. By incorporating stemming, Wordnet Emotion lexicon, unigrams and bigrams on SVM, we derived our best model with accuracy of 50.1475%.

Nevertheless, There is room for future work. First, we assume that the tweets with hashtag in Emotion Intelligence are expressing corresponding emotions. Alternatively, we could try to do crowdsourcing on

Amazon instead of labeling automatically. Second, we'd better find a way to filter different languages and sentences that are within 5 words. Third, we believe there are still ways to reduce irrelevant features such as using PCA and heuristic algorithm. Fourth, we could possibly combine unigrams and bigrams at decision level which may result in a better accuracy. Fifth, it's also helpful to use semi-supervised methods[3] such as boostraping to bring about new features while training. Sixth, We can also make use of other tweets corpus with polarity (Tokuhisa, 2008) and use two step emotion classification to expand from polarity to emotion classification tasks[2].

References

[1] Turney, Peter D. "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews." Proceedings of the 40th annual meeting on association for computational linguistics. Association for Computational Linguistics, 2002.

[2] Tokuhisa, Ryoko, Kentaro Inui, and Yuji Matsumoto. "Emotion classification using massive examples extracted from the web." Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1. Association for Computational Linguistics, 2008.

[3] Dav, D., Oren T., and Ari R.: "Semi-supervised recognition of sarcastic sentences in twitter and amazon." Proceedings of the Fourteenth Conference on Computational Natural Language Learning. Association for Computational Linguistics, 2010.

[4] B. Pang and L. Lee. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1-2): 1-135, 2008.

[5] Go, A., Huang, L., Bhayani, R.: Twitter sentiment classification using distant supervision. In: CS224N Project Report, Stanford (2009)

[6] Liu, Bing. "Sentiment analysis and opinion mining." Synthesis Lectures on Human Language Technologies 5.1 (2012): 1-167.

[7] Mower, Emily, Maja J. Mataric, and Shrikanth Narayanan. "A framework for automatic human emotion classification using emotion profiles." Audio, Speech, and Language Processing, IEEE Transactions on 19.5 (2011): 1057-1070.

[8] Dav, D., Oren T., and Ari R. "Enhanced sentiment learning using twitter hashtags and smileys." Proceedings of the 23rd International Conference on Computational Linguistics: Posters. Association for Computational Linguistics, 2010.

[9] T. Wilson, J. Wiebe, and P. Homann. Recognizing contextual polarity in phrase-level sentiment analysis. In Proceedings of Human Language Technologies Conference/Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP 2005), Vancouver, CA, 2005. ctionReference

[10] Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. "Thumbs up?: sentiment classification using machine learning techniques." Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10. Association for Computational Linguistics, 2002.

[11] D. O. Computer, C. wei Hsu, C. chung Chang, and C. jen Lin. A practical guide to support vector classification chih-wei hsu, chih-chung chang, and chih-jen lin. Technical report, 2003.

[12] J. Read. Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In Proceedings of ACL-05, 43nd Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2005.

[13] Barbosa, Luciano, and Junlan Feng. "Robust sentiment detection on twitter from biased and noisy data." Proceedings of the 23rd International Conference on Computational Linguistics: Posters. Association for Computational Linguistics, 2010.

[14] Pak, Alexander, and Patrick Paroubek. "Twitter as a corpus for sentiment analysis and opinion mining." Proceedings of LREC. Vol. 2010. 2010.

[15] Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python." The Journal of Machine Learning Research 12 (2011): 2825-2830.

[16] Chang, Chih-Chung, and Chih-Jen Lin. "LIBSVM: a library for support vector machines." ACM Transactions on Intelligent Systems and Technology (TIST) 2.3 (2011): 27.

[17] Loper, Edward, and Steven Bird. "NLTK: The natural language toolkit." Proceedings of the ACL-02 Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics-Volume 1. Association for Computational Linguistics, 2002.

[18] Miller, George A. "WordNet: a lexical database for English." Communications of the ACM 38.11 (1995): 39-41.

[19] Makice, Kevin. Twitter API: Up and running: Learn how to build applications with the Twitter API. O'Reilly Media, 2009.

[20] Kouloumpis, Efthymios, Theresa Wilson, and Johanna Moore. "Twitter sentiment analysis: The good the bad and the omg." Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media. 2011.

Appendix: More Examples

RT @BlackaneseKelly: 5 more weeks of school left.👏🏼👌🏼	happy
RT @versoshits: Seas como seas, eres hermosa. http://t.co/LUZqhY9Zj3	sad
RT @Treaseyy: High school really does change you, it's like ion even know acomfs anymore.. They let that "oh you pull all the hoes" get to ...	sad
RT @LifeAsBros: Couples at my school. http://t.co/vmO6l6sNl5	happy
'Rolf hariss went to Rob's school just after he became famous' 'D'you reckon that's what turned Rob gay' LOOOOOOOOL	ashamed
RT @RelatableQuote: This school year is coming to an end 🙌🏼http://t.co/LjM3T5Vjr9	sad
Blood drives = no school for meeee	angry
#ImSoTiredOf 1. People 2. Slow Wi-Fi 3. Drama 4. Homework 5. Exam 6. Being tired 7. Being ignored 8. Sunday 9. School	angry
"@munirahhashix: No school tomorrow again hopefully 😫" you had no school today?	happy
Didnt go to school today I felt so sick lastnight & this morning <<	sad
RT @Chistefavoritos: —Pikachu, ve a agarrar ese pollito. —¡PIKA, PIKA! —No pica, no seas marica. ¡¡HACE PÍO!!	sad
RT @steena_tedesco: Walking up the stairs at school is honestly such a challenge 😳	angry
RT @zarrythrusts: i didn't want to be on twitter during school today but louis has a new tattoo?????????	angry
Let me qo qet mi sister outa school	angry
So much for a diet.... My mom takes me for Chinese food after school😒	sad
Damn im tired of school and I just got here.	angry
Dear school, we're not machines.	ashamed
Skipped school today skipping school tomorrow haha	happy
RT @RelatableQuote: This school year is coming to an end 🙌🏼http://t.co/LjM3T5Vjr9	sad
Photoset: aarontvelts: "what do you do?" "I'm a teacher" "Really? So is my brother! What school?" "It's uh a... http://t.co/lL8Olzm7Qj	angry
I went to school for maybe not even an hour & then left idk	angry
RT @RelatableQuote: This school year is coming to an end 🙌🏼http://t.co/LjM3T5Vjr9	sad
Old School Subprime Slowly Returns: A mortgage lender operating from California is providing home loan financi... http://t.co/cLatzliGU7	angry
@smittydoes currently watching School of Rock, so no you're not alone	angry
@taramrich @dez_norris @rwilliams629 @tmcotney @spiazza505 @bryna_lamb I don't know what you're talking about I love school and Mondays	angry
Kenwood High School Varsity Lacrosse Travis Manion Foundation night. Truly inspiring. "If Not Me, Then Who..." http://t.co/IfpqTNqLF1	sad
RT @xSmiley_Guwopp: Yea I'm gone Really Need Some ice cream After School .	sad
"@Rol_Lex: @shan_chrme_rain I'm hungry and stuck at school 😞FedEx me some food please 😁" <<< I still owe u dinner in EP!! Lol	angry
I sweaaa i neva did like school boys behh	ashamed

As can be seen from the examples, many of the tweets are predicting well. This further illustrates the effectiveness for our framework.