Affect Analysis and Classification on Twitter

Yayang Tian, Tao Feng, Chun Chen

University of Pennsylvania

Computer and Information Science

April 30, 2013

Significance

"We try to classify Tweets into **one of the six affects** which is seldom studied before."

Related Work

1. Broad overview for opining mining (Pang & Lee, 2008)
2. Microblog with SVM and CRF (Yang et al., 2007)
3. Emoticons like ":-)" (Read, 2005)
4. Thumb up and Thumb down (Go et al., 2009)

Formulation

1. How to collect our data? Use existing ones or our own?
2. How to label our data? Manually or Automatically?
3. Which features are suitable for Tweets?
4. What machine learning methods are we going to use?

Contribution of Our Work

1. Presented a pilot study for six-class affect classification on Twitter.
2. Presented a method for automatically collecting Tweets affect corpus.
3. Created a large Tweets corpus consisting of six emotions: "happy, sad, angry, afraid, ashamed, neutral".
4. Build a six-class classification system for microblogging. Adopted various features for tweets and three standard classifier. Outperform baseline approach by 21.197%.
5. Exposed the importance for stemming, affect dictionary, smiley, information gain, and SVM in six-class classification.
6. Conducted experimental evaluations on real-time Tweets which demonstrates the effectiveness of our system.
7. Build a web application that can classify and summarize emotions on Twitter in real-time.

Corpus Creation

We crawled 42416 Tweets using Twitter Streaming API:

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

Table : Data Collection

Labeling

Labeling the Training Data

1. We did not annotate by ourselves.
2. We collect Tweets based on words in emotion intelligence.
3. As long as there is one word from Tweet appears in the dict, we add it to training data.

Experimental Baseline:

Baseline:

1. Our baseline: Naive Bayes classification on unigrams for raw data
2. Optional baseline: count of the words appearing SemEval

Preprocessing:

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

Table : Preprocessing and significance

Examples after Preprocessing

Raw Tweets
Msdebramaye I heard about that contest! Congrats girl!
Good night #Twitter and #TheLegionoftheFallen. 5:45am cimes awfully ea
Just had some bloodwork done. My arm hurts
Processed Tweets
AT_USER i heard about that contest! congrats girl!!
good night twitter and thelegionofthefallen. 5:45am cimes awfully early!
just had some bloodwork done. my arm hurts

Table : Preprocessing Example

Unigram

We tried different methods based on Unigrams features for SVM classifier:

Method	Accuracy
Raw data with no preprocessing	46.4602%
Include icon like smile icon	46.397%
Include emoticons like :) : (	46.7446%
Include Stemmer using Lancaster	49.0624%
Word only without punctuations	43.0784%
add emoticons with stemmer and emoticons	49.3257%
Add Stoplist with stemmer and emoticons	49.3679%

Table : Predictions based on on SVM-Unigrams

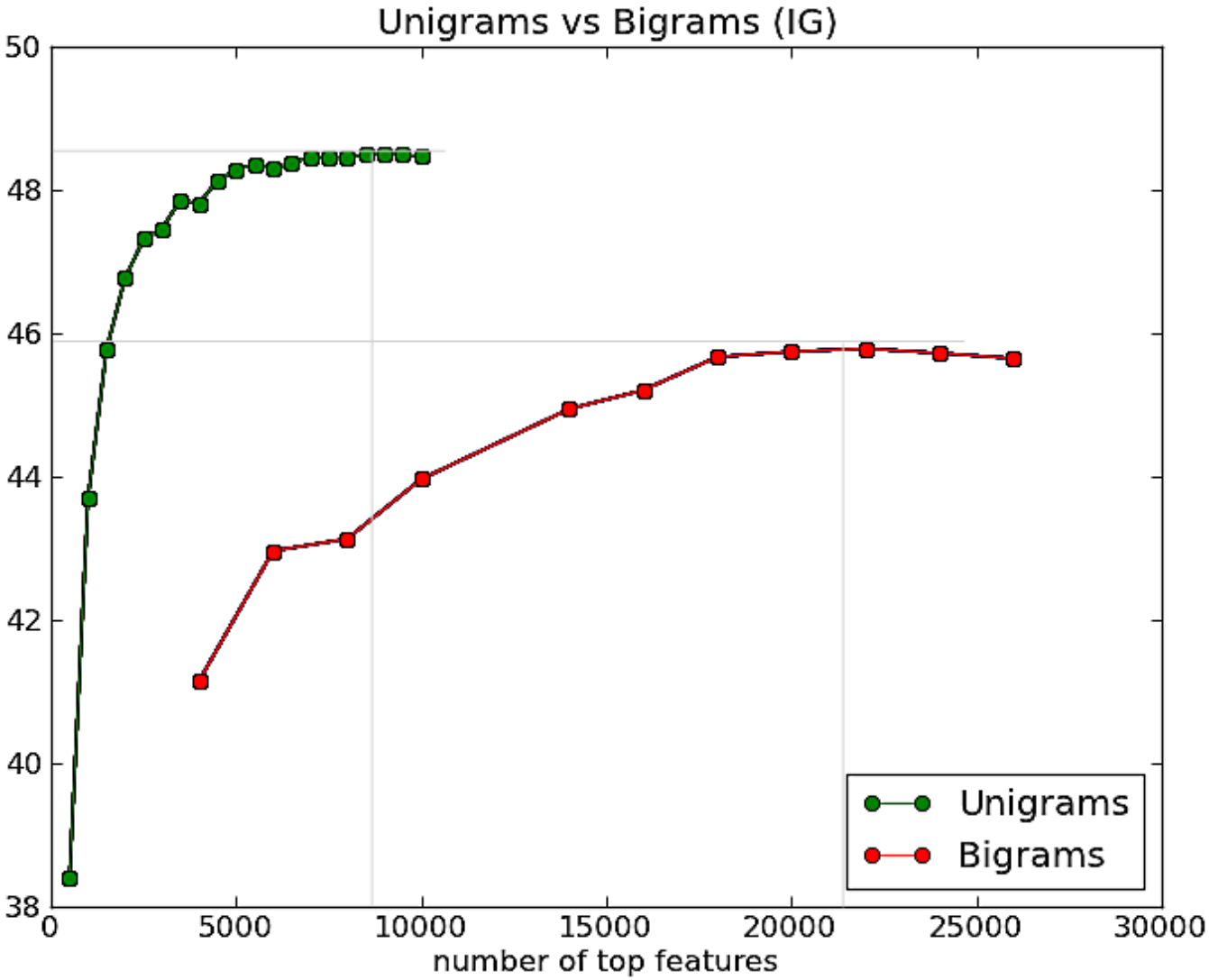
Information Gain

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%

Table : SVM prediction acc for IG

Information Gain

Compare between unigrams and bigrams:



Information Gain: Unigrams & Bigrams

Unigrams Plus Bigrams

1. Mix 4500 Unigrams and 8000 Bigrams: Accuracy 49.4732%
2. Mix 8000 Unigrams and 16000 Bigrams: Accuracy 49.9157%

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

Table : SVM Prediction using IG

Conclusion

Best performance are given by combing unigrams and bigrams.

Affect Dictionaries

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 %

Table : SVM Prediction based on Dictionaires

Conlusion:

1. Emotion dictionaries help improved by 3%
2. WordNet has best indication of emotion

Emotion Profile

Steps

1. Train five libSVM binary classifier. Like "Happy" or "Not Happy"
2. Pick the one with max confidence/hyperplane distance as prediction

Classifier	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%

Table : Five B-Classfiers for EP

EP Result

Total: 9492; Correct: 4509.0; **Accuracy: 47.5031605563%**

Different Machine Learning Methods

Classifier	Accuracy
Naive Bayes	28.9506%
Max Entropy	39.3234%
SVM	50.1475%

Table : Three Standard Classifers

Conclusion

Among them, SVM has best performance for our data.

Summary of Best Model

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

Table : Summary for each step

Confusion Matrix of Best Model

Accuracy = 50.1475% (4760/9492) (classification)
Confusion Matrix for our best result:
0: 0: 854, 1: 209, 2: 420, 3: 94, 4: 151,
1: 0: 47, 1: 1236, 2: 121, 3: 80, 4: 103,
2: 0: 412, 1: 535, 2: 1150, 3: 239, 4: 193,
3: 0: 299, 1: 413, 2: 336, 3: 639, 4: 143,
4: 0: 180, 1: 394, 2: 262, 3: 101, 4: 881

Web Application

1. Enter search query.
2. Return most relevant and recent tweets around you.
3. Predict their emotions into one of the six categories.

Web Application: Example

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

Web Application: More Tweets

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

Web Application: More Tweets

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

Future Work

Refine Corpus:

1. Need annotation before training.
   Accuracy for human even lower than our classification system (42.500% versus 50.14750%).
2. Eliminate different languages and short sentences.

Refine features:

1. PCA and heuristic algorithms to reduce features.
2. Make use of #hashtag for unigram features of training.

Refine machine learning:

1. Try other Unsupervised methods.
2. Try more kernels for SVM.

Experiment

1. Try to calculate inter-annotation agreement before training.
2. Try to measure correlation and precision/recall.

Conclusion

1. Presented a pilot study for **six-class affect classification** on **Twitter**.
2. Presented a method for **automatically collecting** Tweets affect corpus.
3. Created a large **Tweets corpus** consisting of six emotions: "happy, sad, angry, afraid, ashamed, neutral".
4. Build a **six-class classification system** for microblogging. Adopted various features for tweets and three standard classifier. Outperform baseline approach by 21.197%.
5. Exposed the importance for **stemming, affect dictionary, smiley, information gain, and SVM** in six-class classification.
6. Conducted **experimental evaluations on real-time** Tweets which demonstrates the effectiveness of our system.
7. Build a **web application** that can classify and summarize emotions on Twitter in real-time.

The End