

Twitter Sentiment Analysis

Jayant Sharma

Aniruddh Vyas

Mentor – Prof. Amitabha Mukerjee

Motivation

- ▶ Huge Traffic: > 50 million tweets a day; character limit
- ▶ “..microscopic instantiations of mood..”

72 FOLLOWING

19 FOLLOWERS

Tweets

 **Jayant Sharma** @bagghu 2m
delightful!!! : Milan ousting Barca in CL ---> goal.com/en-india/news/
...

 **Jayant Sharma** @bagghu 4m
feeling sad

 **MTV India** @MTVIndia 11 Jul

Method

- Data and Instrument
 - a corpus of 5,156,047 tweets published by Twitter users (time period: Jan 2009 – March 2010)
 - a well established psychometric instrument, the Profile of Mood States

Extended POMS–bi

- ▶ 6 bipolar dimensions of mood:
 - Composed/Anxious
 - Agreeable/Hostile
- ▶ 72 mood adjectives; 12 for each mood dimension:
 - For eg: angry to measure along ‘hostile/agreeable’ mood dimension
- ▶ Extend POMS using WordNet: POMS–bi–ex
 - Eg: angry → wild, raging, tempestuous

Tweet Processing

Try being ANGRY, sad and ... nervous at the same time!!!!

try being angry, sad and nervous at the same time

angry sad nervous time

angri sad nervou time

angri sad nervou time
(checked against POMS-bi-ex lexicon)

(composed, aggreable, elated, confident, tired,
confused)

Mood_Vector: (-1, -1, -1, 0, 0, 0)

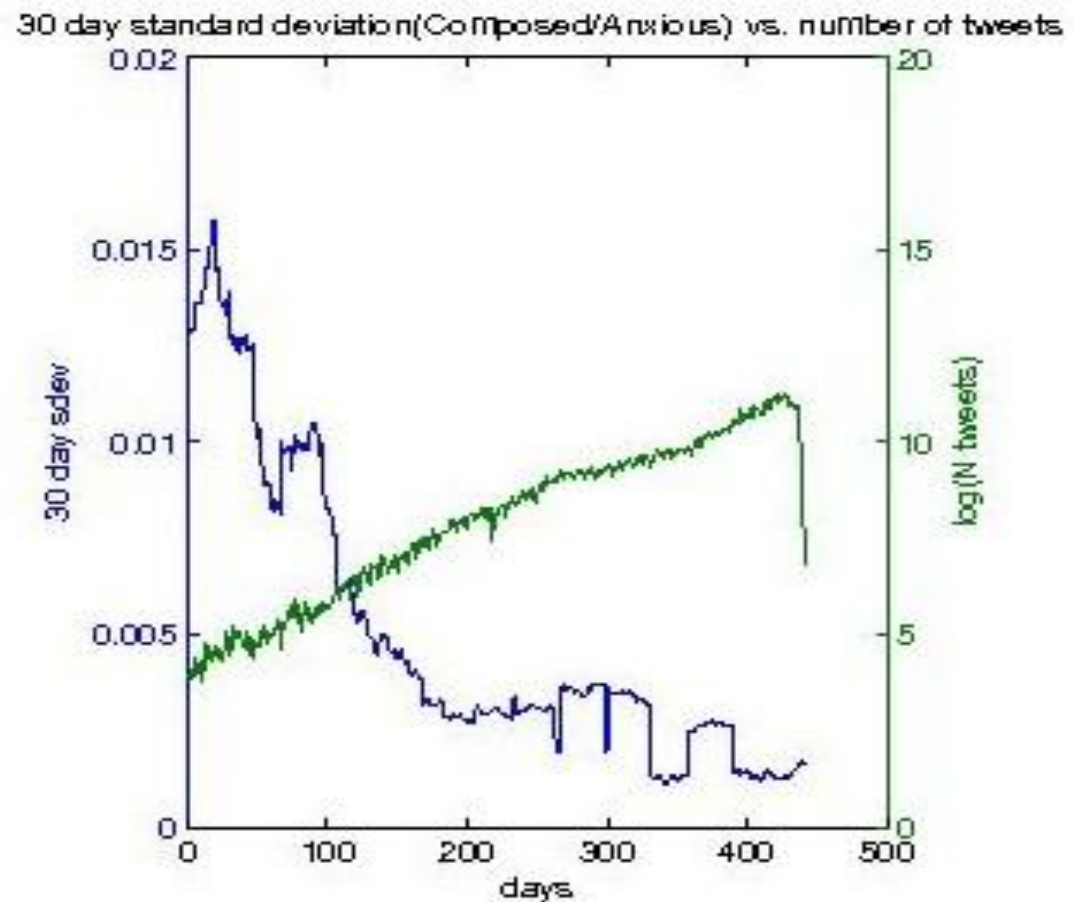
average mood vectors for a date → aggregate
mood vector

Code

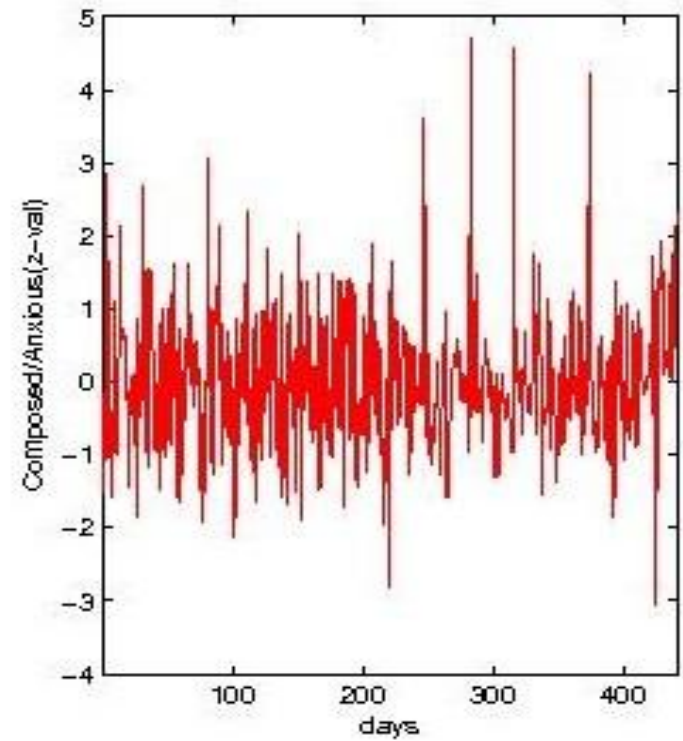
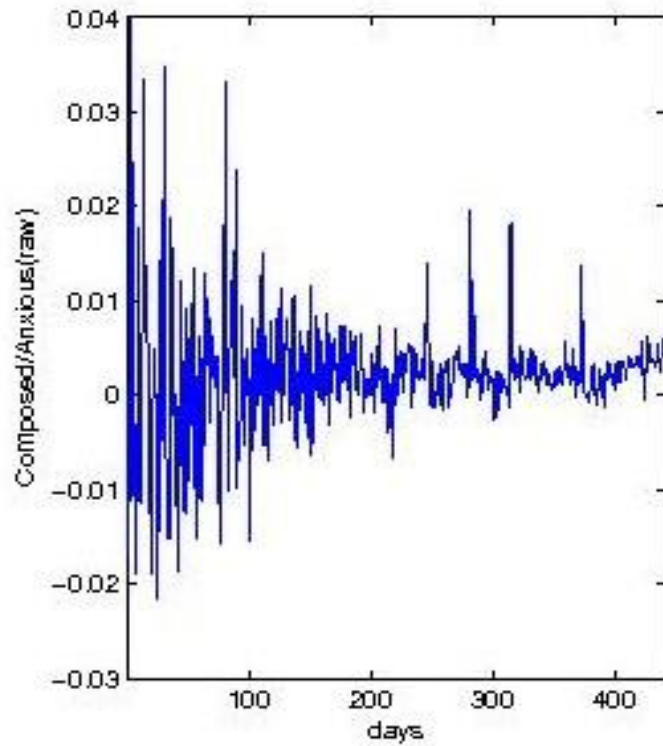
- ▶ PLY 3.4 – a python implementation of lex-yacc
- ▶ porter-stemming.py – python implementation of Martin Porter's stemmer (by Vivake Gupta)

Results

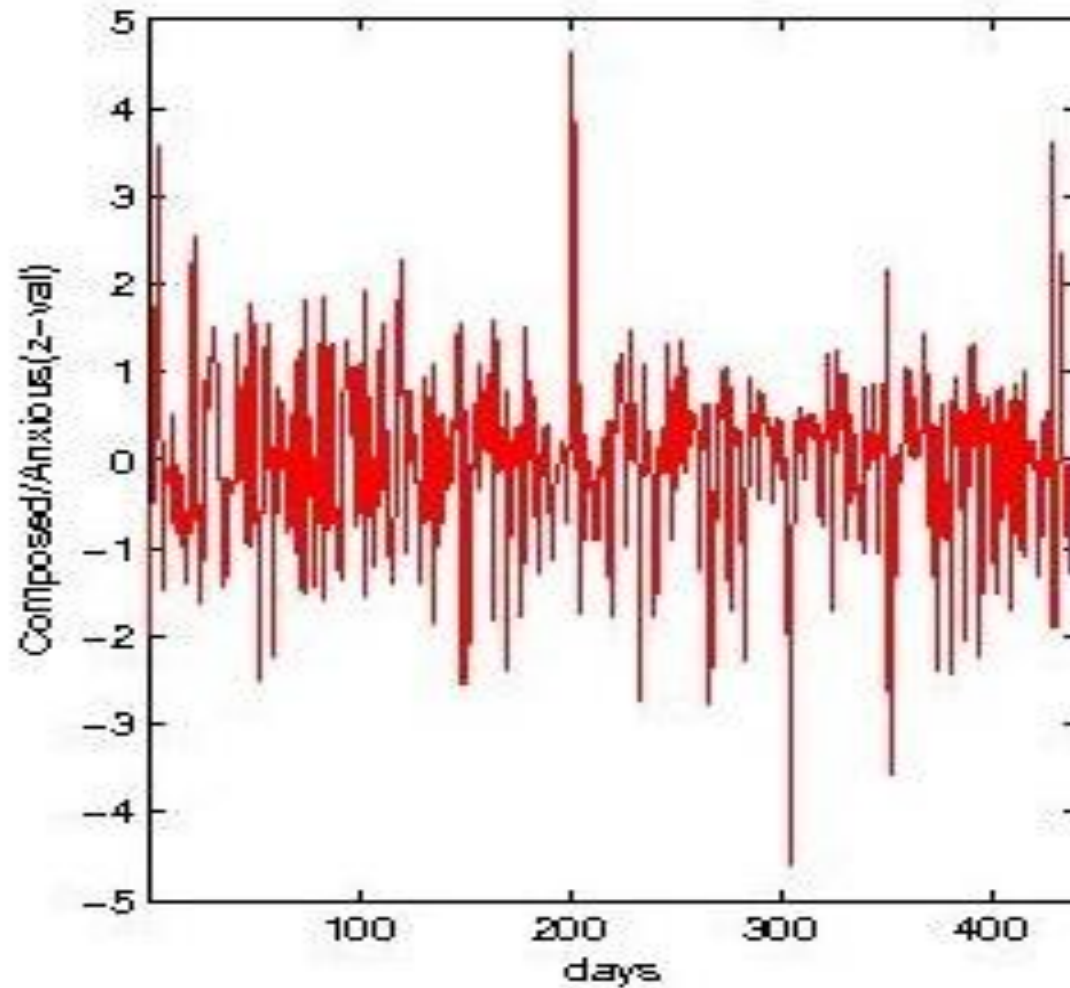
- ▶ Variance increases inversely with number of tweets



conversion to z-scores



Lets eyeball!



Extending our work and alternatives:

- ▶ Expand the POMS lexicon using word co-occurrences, by querying the Web 1T n-gram database
 - ▶ Looking for a correlation between stock market variation and twitter sentiment
- 