Emoticons and Phrases: Status Symbols in Social Media

Simo Tchokni, Diarmuid O Seaghdha Cambridge Uni
Daniele Quercia Yahoo Labs

ICWSM 2014

Introduction

Language study

Social power and language use

Social status (limited to 140 characters)

Contribution

Emphasis on *prediction*

A lexicon of phrases associated with social power

Relationship between *emoticons* and social status

Related Work

first person vs. second person pronoun

politeness and complexity

followers and sentiment

User Prediction metric

#followers: popularity or influence

retweet, mention, Klout

Dataset

258,895 Twitter user and 31.5 M English tweets

Twitter public API between Sep to Dec 2012

seed: popular UK-based news outlets

Dataset

	FOLLOWERS	KLOUT
low cut-off	≤ 87	≤ 16.85
high cut-off	> 1113	> 46.25
Minimum	0	1
Maximum	6,520,279	100
Total **low** users	65,054	43,818
Total **high** users	64,711	43,692
Total users	129,765	87,510
Messages per user	111.6	143.9

Features(n-grams)

unigrams: 2,837,175 and bigrams: 42,296,563

URLs, numbers, punctuation, hashtags, non-ascii

usernames, stock symbols

Features (LIWC)

Dimension	Example words
first person	*I, my, me ...*
second person	*you, your ...*
third person	*she, he, they ...*
cognitive	*believe, choice, apparently ...*
time	*anytime, new, long ...*
past	*arrived, asked, ended ...*
present	*begin, do, want ...*
future	*gonna, may, might ...*
posemo	*nice, outgoing, original ...*
negemo	*no, offend, protest ...*

Features (NRC)

Dimension	Example words
anger	*punch, reject, ruined ...*
anticipation	*punctual, savor, romantic ...*
disgust	*abuse, prejudiced, sickening ...*
fear	*abandon, rifle, scarce ...*
joy	*blessed, romantic, score ...*
negative	*misery, oversight, quit ...*
positive	*mate, nap, plentiful ...*
sadness	*blue, shatter, starvation ...*
surprise	*coincidence, catch, secrecy ...*
trust	*scientific, save, toughness ...*

Features (Emoticons)

Emoticon use: avg #emoticon per tweet

fraction of positive/negative

Emoticons dictionary: 78 positive and 57 negative

Features (length)

average word length

average tweet length

number of large words

average word length greater than 6 or not?

Features (spelling)

fraction of misspelled words

List of common misspelling wikipedia

Features (Punctuation)

fraction of tweets containing at least one ?

fraction of tweets containing at least one !

Features (Elongation)

common phenomenon on Twitter

like cooooooool! shows intensification and informality

fraction of words that a user elongate

Features (mention & RT)

level of engagement

fraction of user's tweets that are retweet

fraction of tweets that are addressed to other users

level of engagement

fraction of user’s tweets that are retweet

fraction of tweets that are addressed to other users

Classification Task

Support Vector Machine and Linear Regression

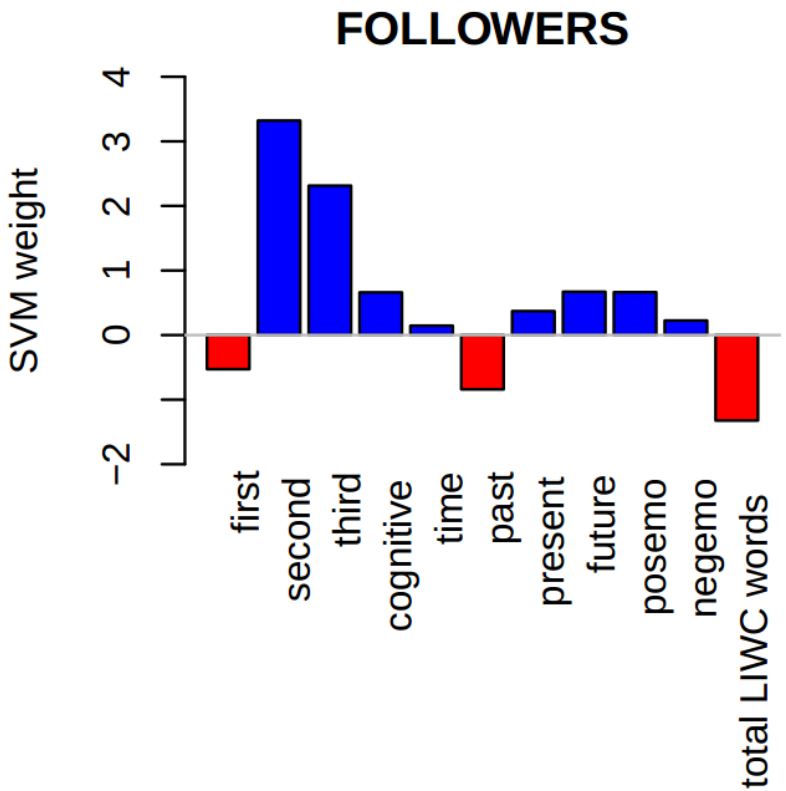
separate classifier for each set of features

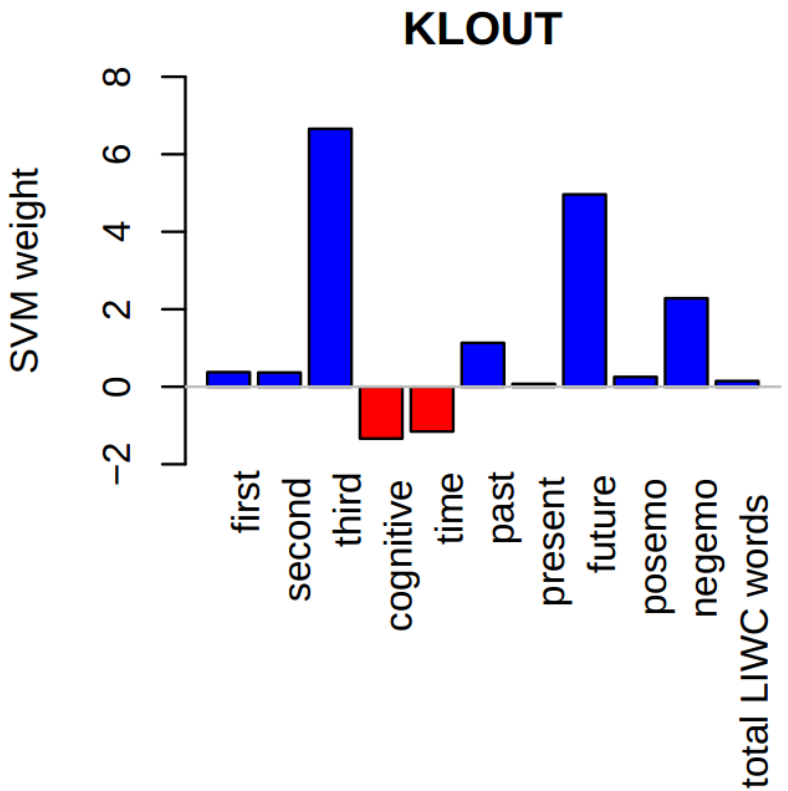
10-fold cross validation

Classification Result

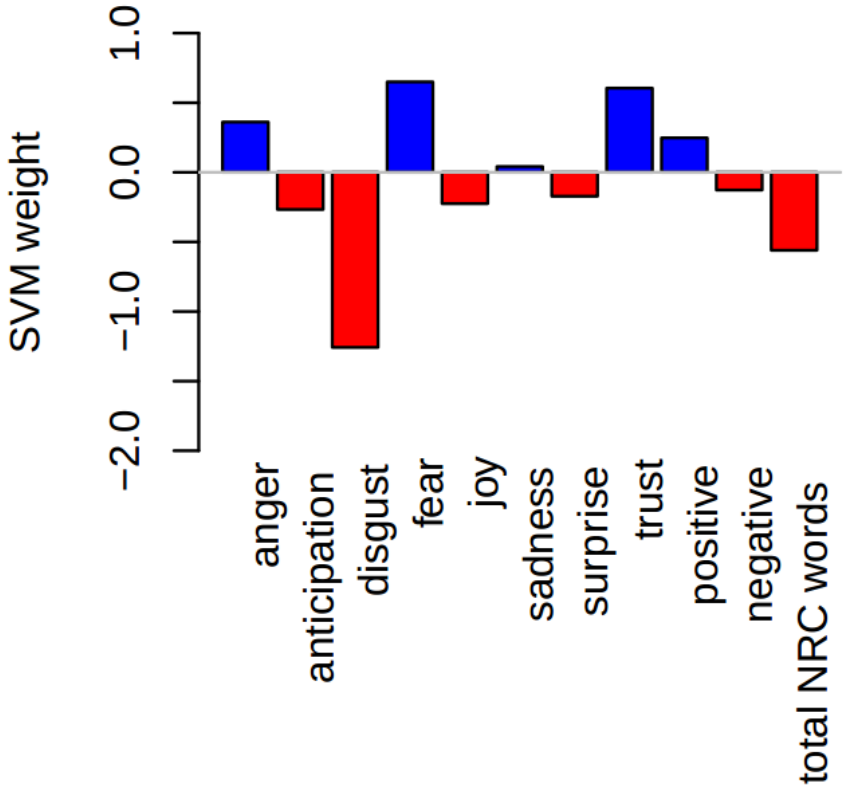
Features Used	FOLLOWERS	KLOUT
Baseline	50.00	50.00
unigrams	81.38***	80.43***
bigrams	80.59***	77.26***
NRC	64.30***	59.95***
LIWC	65.42***	65.11***
emoticons	66.46***	61.06***
tweet and word length	63.17***	58.98***
spelling	48.79	61.67
word elongation	49.02**	50.07**
punctuation	63.53**	54.11**
mentioning others	60.24***	57.95***
retweeting	70.02***	64.87***
All features	**82.37*****	**81.28*****

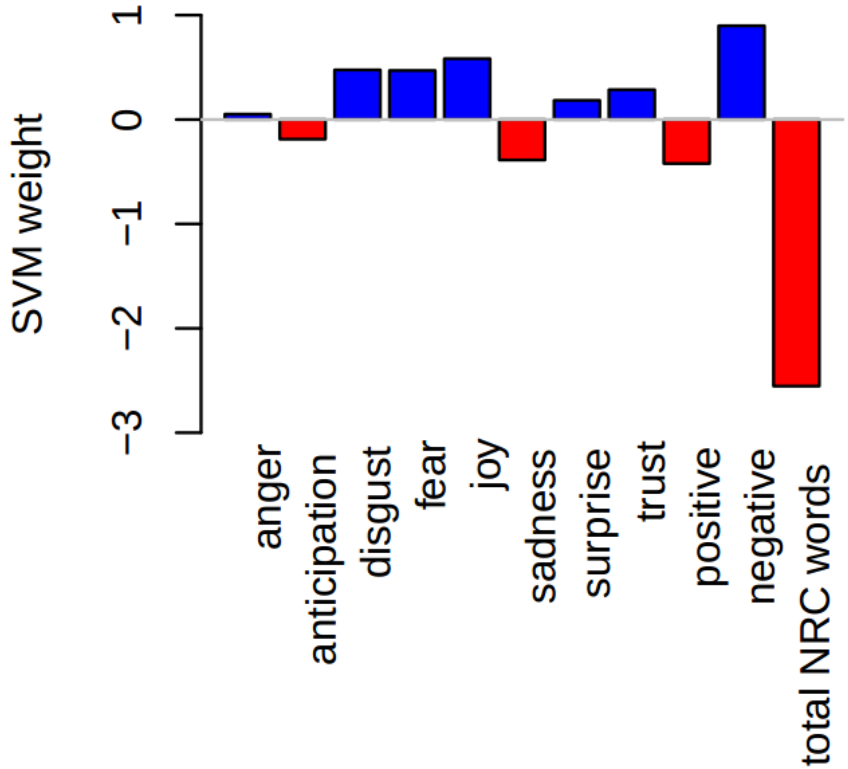
weight vectors (LIWC)



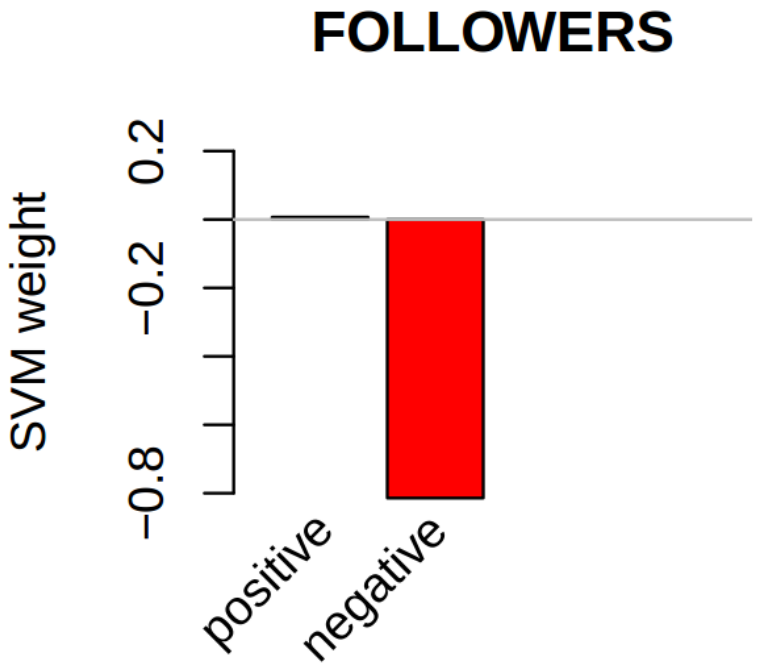


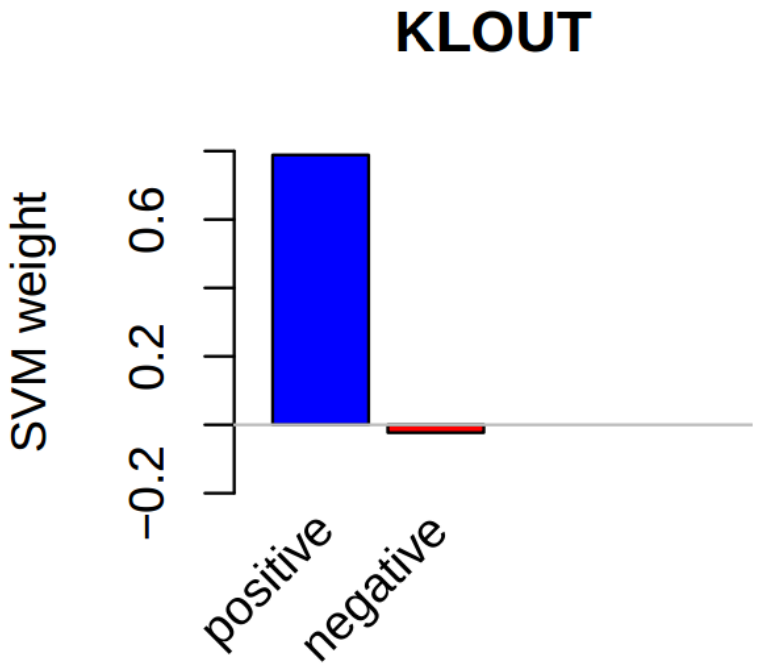
weight vectors (NRC)





weight vectors (Emoticon)





Top unigrams

low	high
surely	rts
shame	backstage
;p	cc
bloody	washington
whilst	questions
uni	nope
cameron	hollywood
wondering	nyc
yg	tells
thinks	dm
gutted	bloggers
babeeee	headed
rubbish	shows
mum	sorry
preparing	toronto
twittering	powerful
debra	y'all
boring	announced
luck	thx

Top bigrams

low	high
avatar now	in la
at work	the rt
well done	rt my
an iphone	:) rt
bring on	headed to
just seen	white house
managed to	rt i
loving the	u s
for following	you're welcome
bank holiday	you missed
roll on	lindsay lohan
the follow	thanks so
oh dear	talks about
come on	w the
you dauntons	rt just
the welcome	thank u
back from	your favorite
the train	in nyc
this space	sorry i

Cross Validation

121,823 Facebook users

5.97 message per user

60% accuracy

Conversation prediction

in a dyadic conversation

predict which of the two users is more popular

Dataset

Nov 2012 to Feb 2013 random sample of users

construct conversation using reply-to information

2,158 conversation between 1,511 users

sample of additional tweets for each user

Features

conversation start: which user started the conversation

deviation: how much x deviates from his usual way of writing when talking to y

Echoing: whether x uses words for the first time after y has used it

Classification Task

Support Vector Machine

10-fold cross validation

different classifiers per features

Classification Result

	Feature Set	Accuracy
	Baseline	50.00
(1)	style deviation	56.88**
(2)	emotion deviation	53.68**
(3)	word choice deviation	**71.56****
(4)	style echoing	48.96*
(5)	emotion echoing	50.07*
(6)	word choice echoing	49.28
(7)	conversation start	58.27***
(8)	unigrams	53.96*
(9)	NRC	51.64
(10)	LIWC	50.35
(11)	emoticons	49.98
(12)	tweet and word length	53.50
(13)	spelling	49.70
(14)	word elongation	48.49
(15)	punctuation	47.34
(16)	All features	71.33***

Classification Result

conversation prediction is a difficult task

deviation is a good feature but echoing not!