oiioiioii oxford
oiioiioii internet institute
oiioiioii university of oxford

ÅBO AKADEMI
Information Studies

Introduction to Sentiment strength detection

Mike Thelwall
University of Wolverhampton, UK

Statistical Cybermetrics Research Group

CYBEREMOTIONS

UNIVERSITY OF WOLVERHAMPTON

Sentiment Strength Detection with *SentiStrength*

1. Detect positive and negative sentiment *strength* in short informal text
    1. Develop workarounds for lack of standard grammar and spelling
    2. Harness emotion expression forms unique to MySpace or CMC (e.g., :-) or haaappppyyy!!!)
    3. Classify simultaneously as positive 1-5 AND negative 1-5 sentiment
2. Apply to MySpace comments and social issues

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12), 2544-2558.

SentiStrength 1 Algorithm - Core

- List of 890 positive and negative sentiment terms and strengths (1 to 5), e.g.
  - ache = -2, dislike = -3, hate=-4, excruciating -5
  - encourage = 2, coolest = 3, lover = 4
- Sentiment strength is highest in sentence; or highest sentence if multiple sentences

Examples

positive, negative

-2

- My legs ache. 1, -2

3

- You are the coolest. 3, -1

- I hate Paul but encourage him. 2, -4

-4 2

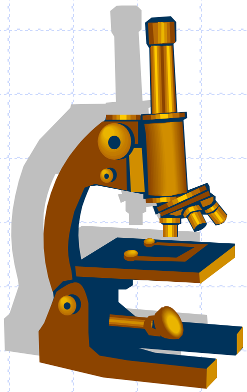
Term Strength Optimisation

- Term strengths (e.g., ache = -2) initially fixed by a human coder
- Term strengths optimised on training set with 10-fold cross-validation
  - Adjust term strengths to give best training set results then evaluate on test set
  - E.g., training set: “My legs <u>ache</u>”: coder sentiment = 1,-3 => adjust sentiment of “ache” from -2 to -3.

Summary of sentiment methods

- **sentiment word strength list** terrify=-4
  - “**miss**” = +2, -2
- **spelling corrected** nicce -> nice
- **booster words** alter strength **very** happy
- **negating words** flip emotions **not** nice
- **repeated letters** boost sentiment/+ve niiiice
- **emoticon list** :) =+2
- **exclamation marks** count as +2 unless –ve hi!
- **repeated punctuation** boosts sentiment good!!!
- **negative emotion ignored in questions** u h8 me?

Experiments



- Development data = 2600 MySpace comments coded by 1 coder
- Test data = 1041 MySpace comments coded by 3 independent coders
- Comparison against a range of standard machine learning algorithms

Test data: Inter-coder agreement

Krippendorff's inter-coder weighted alpha = 0.5743 for positive and 0.5634 for negative sentiment

Only moderate agreement between coders
but it is a hard 5-category task

Comparison for 1041 MySpace texts	**+ve agree-ment**	**-ve agree-ment**
Coder 1 vs. 2	**51.0%**	**67.3%**
Coder 1 vs. 3	**55.7%**	**76.3%**
Coder 2 vs. 3	**61.4%**	**68.2%**

Results:+ve sentiment strength

Algorithm	**Optimal #features**	**Accuracy**	**Accuracy +/- 1 class**	**Correlation**
SentiStrength	**-**	**60.6%**	**96.9%**	**.599**
Simple logistic regression	700	**58.5%**	96.1%	**.557**
SVM (SMO)	800	**57.6%**	**95.4%**	**.538**
J48 classification tree	700	**55.2%**	**95.9%**	.548
JRip rule-based classifier	700	**54.3%**	96.4%	**.476**
SVM regression (SMO)	100	**54.1%**	97.3%	**.469**
AdaBoost	100	**53.3%**	**97.5%**	**.464**
Decision table	200	**53.3%**	96.7%	**.431**
Multilayer Perceptron	100	**50.0%**	*94.1%*	**.422**
Naïve Bayes	100	**49.1%**	**91.4%**	**.567**
Baseline	-	**47.3%**	**94.0%**	**-**
Random	-	**19.8%**	**56.9%**	**.016**

Results:-ve sentiment strength

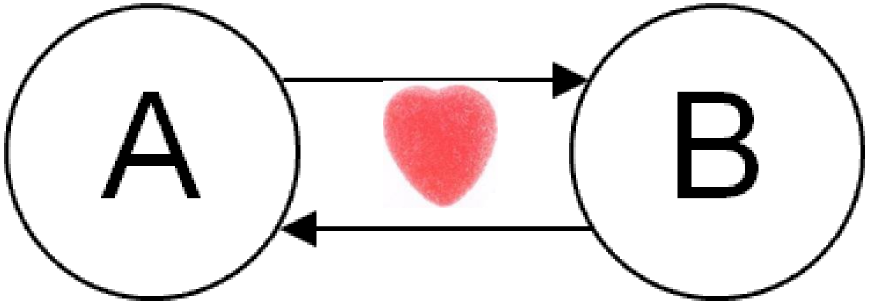
Algorithm	Optimal #features	Accuracy	Accuracy +/- 1 class	Correlation
SVM (SMO)	100	**73.5%**	**92.7%**	**.421**
SVM regression (SMO)	300	**73.2%**	**91.9%**	**.363**
Simple logistic regression	800	**72.9%**	**92.2%**	**.364**
SentiStrength	-	**72.8%**	**95.1%**	**.564**
Decision table	100	**72.7%**	**92.1%**	**.346**
JRip rule-based classifier	500	**72.2%**	**91.5%**	**.309**
J48 classification tree	400	**71.1%**	**91.6%**	**.235**
Multilayer Perceptron	100	**70.1%**	**92.5%**	**.346**
AdaBoost	100	**69.9%**	**90.6%**	-
Baseline	-	**69.9%**	**90.6%**	-
Naïve Bayes	200	**68.0%**	**89.8%**	**.311**
Random	-	**20.5%**	**46.0%**	**.010**

Example differences/errors

- *THINK* 4 THE ADD
  - Computer (1,-1), Human (2,-1)
- 0MG 0MG 0MG 0MG 0MG 0MG 0MG 0MG!!!!!!!!!!!!!!!!!!!!N33N3R!!!!!!!!!!!!!!!
  - Computer (2,-1), Human (5,-1)

Application - Evidence of emotion homophily in MySpace

- Automatic analysis of sentiment in 2 million comments exchanged between MySpace friends
- Correlation of 0.227 for +ve emotion strength and 0.254 for –ve
- People tend to use similar but not identical levels of emotion to their friends in messages

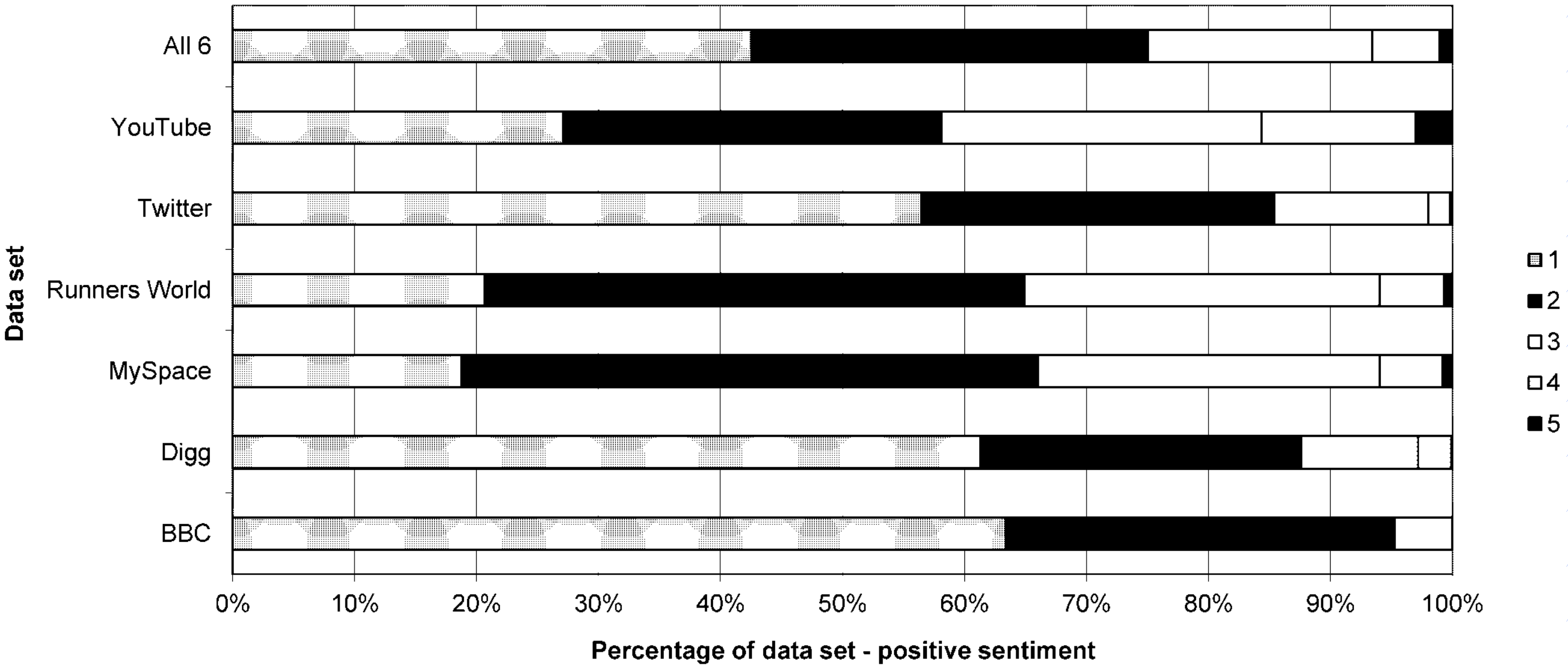


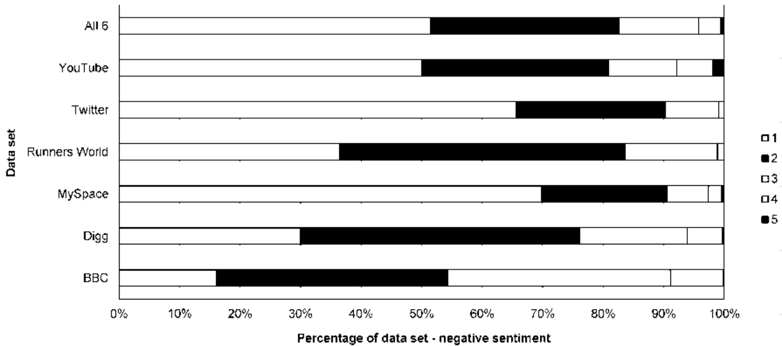
SentiStrength 2

- Sentiment analysis programs are typically domain-dependant
- SentiStrength is designed to be quite generic
  - Does not pick up domain-specific non-sentiment terms, e.g., G3
- SentiStrength 2.0 has extended negative sentiment dictionary
  - In response to weakness for negative sentiment

Thelwall, M., Buckley, K., Paltoglou, G. (submitted).
High Face Validity Sentiment Strength Detection for the Social Web

6 Social web data sets





To test on a wide range of different Social Web text

SentiStrength 2 (unsupervised) tests

Social web sentiment analysis is less domain dependant than reviews

Data set	Positive Correlation	Negative Correlation
YouTube	0.589	0.521
MySpace	0.647	0.599
Twitter	0.541	0.499
Sports forum	0.567	0.541
Digg.com news	0.352	0.552
BBC forums	0.296	0.591
All 6	0.556	0.565

Why the bad results for BBC?

- Long texts, mainly negative, expressive language used, e.g.,
  - David Cameron must be very happy that I have lost my job.
  - It is really interesting that David Cameron and most of his ministers are millionaires.
  - Your argument is a joke.

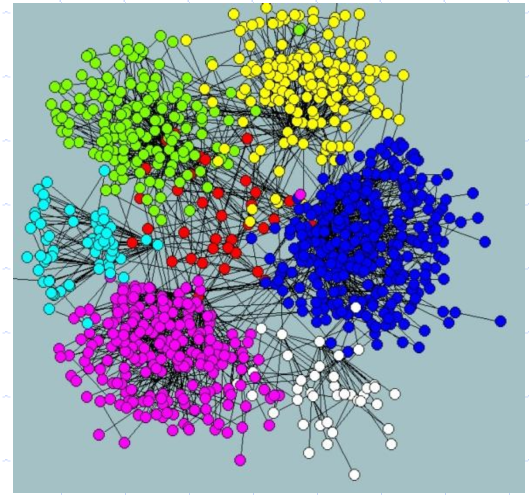
SentiStrength vs. Machine learning for Social Web texts

- Machine learning performs a bit better overall (7 out of 12 data sets/+ve or negative)
  - Logistic Regression with trigrams, including punctuation and emoticons; 200 features
- But has “domain transfer” and “face validity” problems for ***some*** tasks

SentiStrength software

- Versions: Windows, Java, live online (sentistrength.wlv.ac.uk)
- German version (Hannes Pirker)
- Variants: Binary (positive/negative), trinary (positive/neutral/negative) and scale (-4 to +4)
- Sold commercially - & purchasers converting to French, Spanish, Portuguese, & ?

CYBEREMOTIONS



Sentistrength

Collective Emotions in Cyberspace

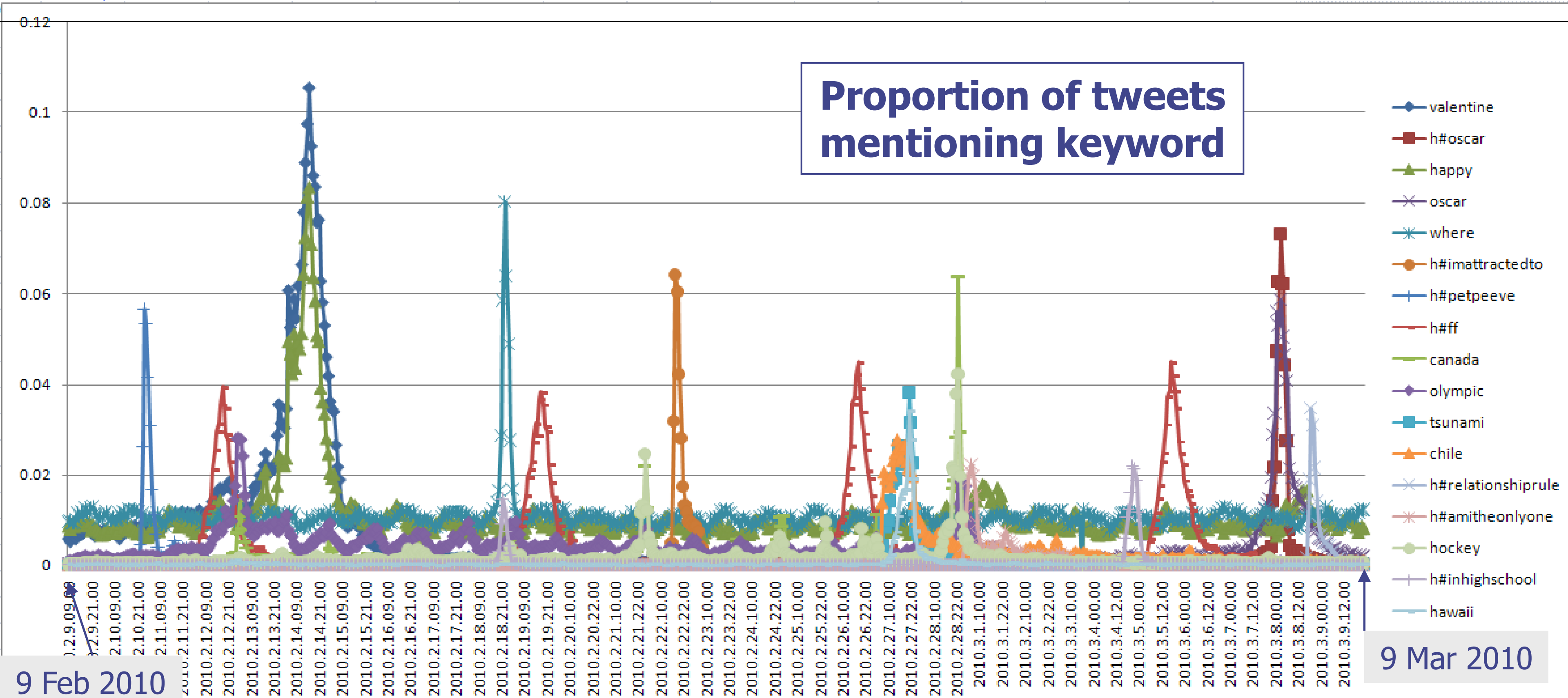
CYBEREMOTIONS = data gathering + **complex systems methods** + ICT outputs

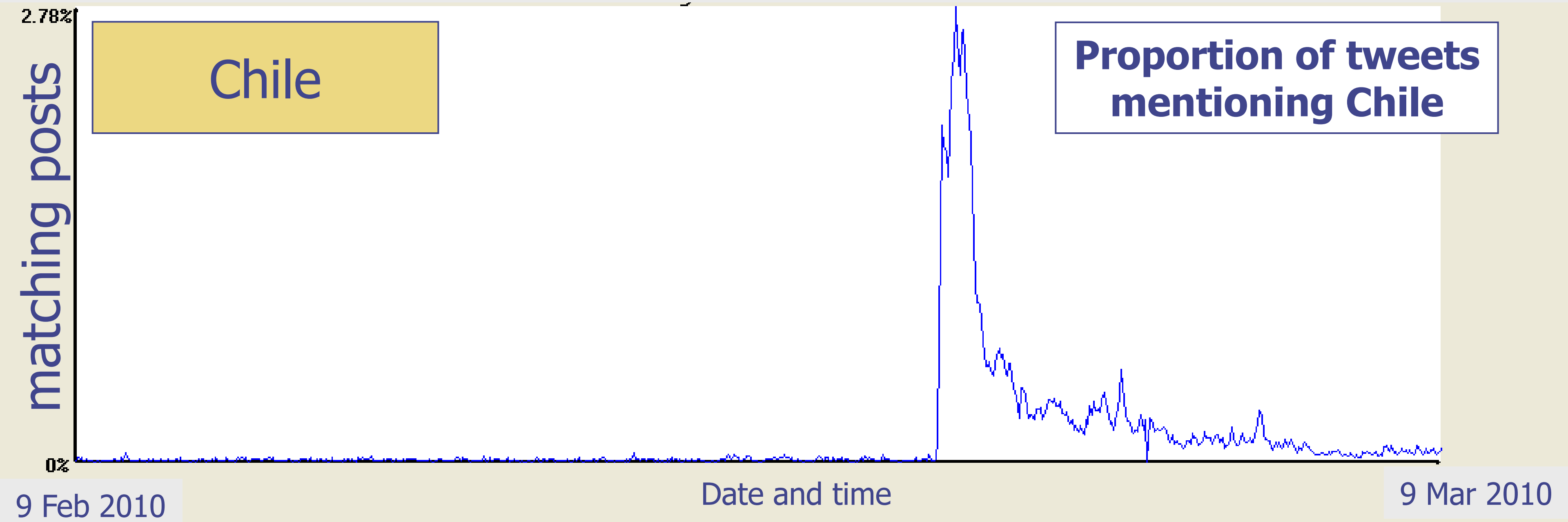
Application – sentiment in Twitter events

- Analysis of a corpus of 1 month of English Twitter posts
- Automatic detection of spikes (events)
- Sentiment strength classification of all posts
- Assessment of whether sentiment strength increases during important events

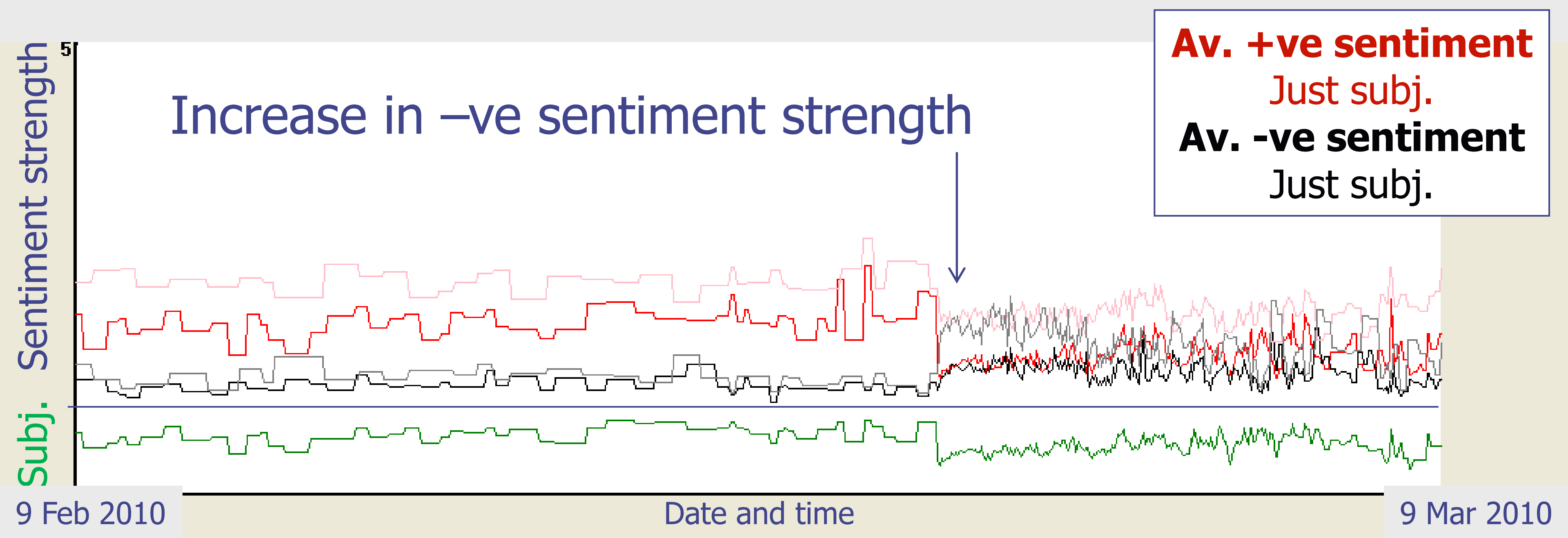
Thelwall, M., Buckley, K., & Paltoglou, G. (2011). Sentiment in Twitter events. *Journal of the American Society for Information Science and Technology,* 62(2), 406-418.

Automatically-identified Twitter spikes

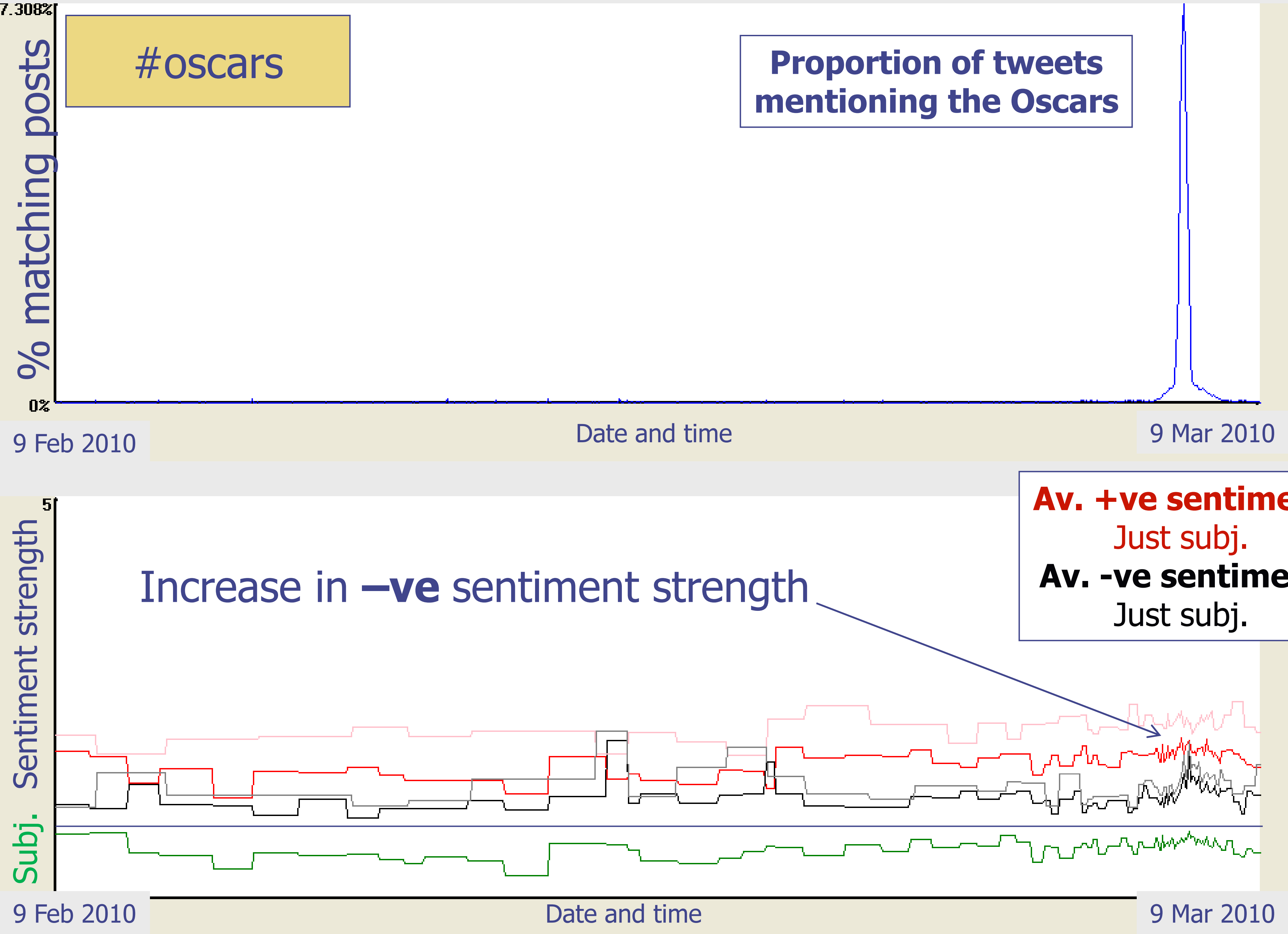




Chile
Proportion of tweets mentioning Chile
matching posts
2.78%
0%
9 Feb 2010
Date and time
9 Mar 2010



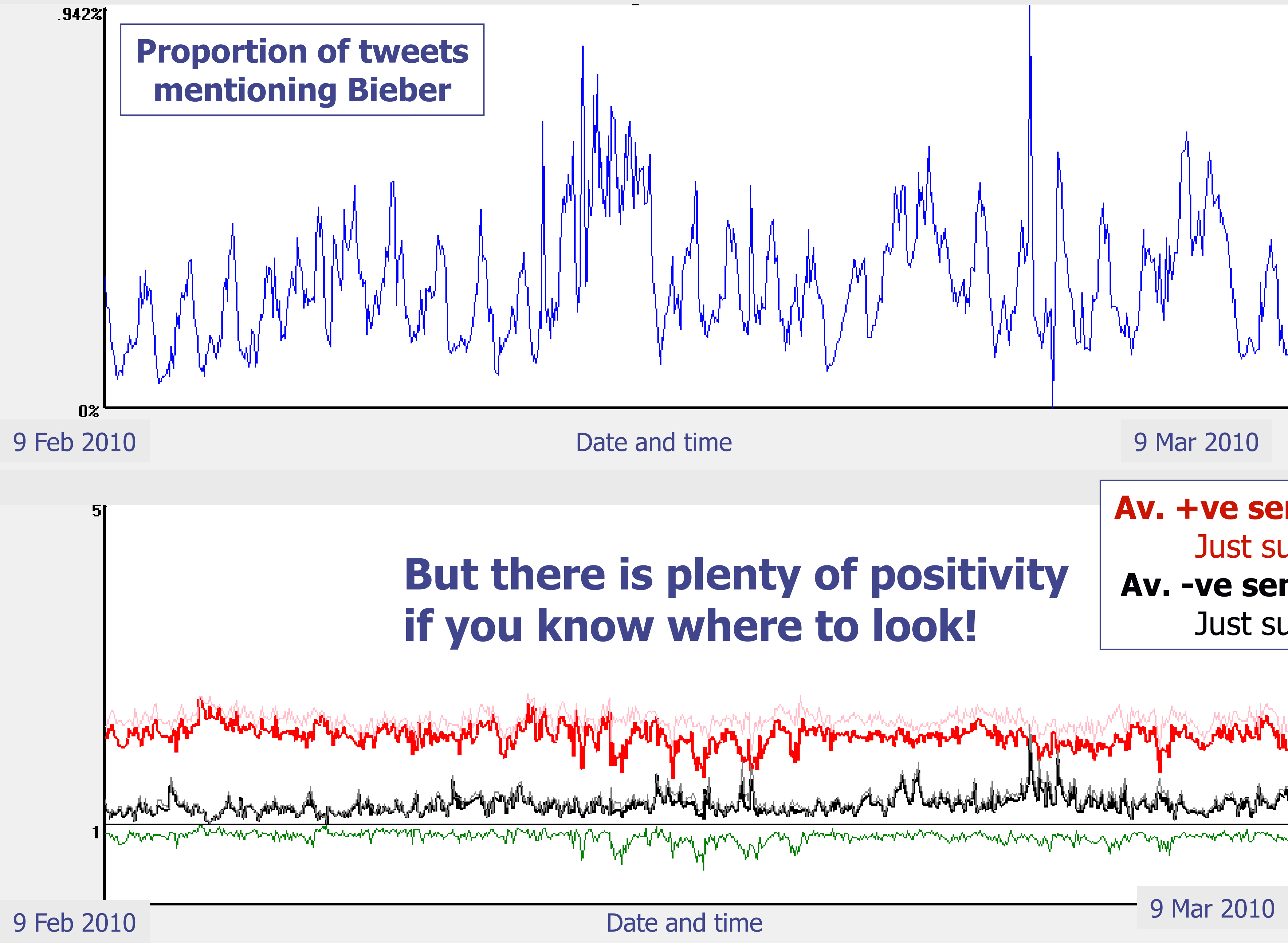
Increase in –ve sentiment strength
Av. +ve sentiment
Just subj.
Av. -ve sentiment
Just subj.
Sentiment strength
Subj.
5
9 Feb 2010
Date and time
9 Mar 2010


#oscars
Proportion of tweets mentioning the Oscars
7.308%
0%
% matching posts
9 Feb 2010
Date and time
9 Mar 2010
5
Sentiment strength
Subj.
Increase in –ve sentiment strength
Av. +ve sentime
Just subj.
Av. -ve sentime
Just subj.
9 Feb 2010
Date and time
9 Mar 2010

Sentiment and spikes

- Analysis of top 30 spiking events
  - Strong evidence ($p=0.001$) that *higher volume hours have stronger negative sentiment than lower volume hours*
  - *Insufficient evidence* ($p=0.014$) that higher volume hours have different positive sentiment strength than lower volume hours

=> Spikes are typified by increases in *negativity*


Proportion of tweets mentioning Bieber
.942%
0%
9 Feb 2010
Date and time
9 Mar 2010
But there is plenty of positivity if you know where to look!
Av. +ve se
Just s
Av. -ve se
Just s
5
1
9 Feb 2010
Date and time
9 Mar 2010

YouTube Video comments

- Short text messages left for a video by viewers
- Up to 1000 per video accessible via the YouTube API
- A good source of social web text data

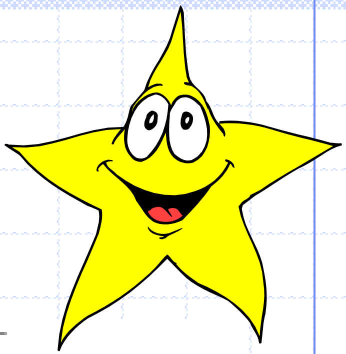
THAT'S RIGHT.

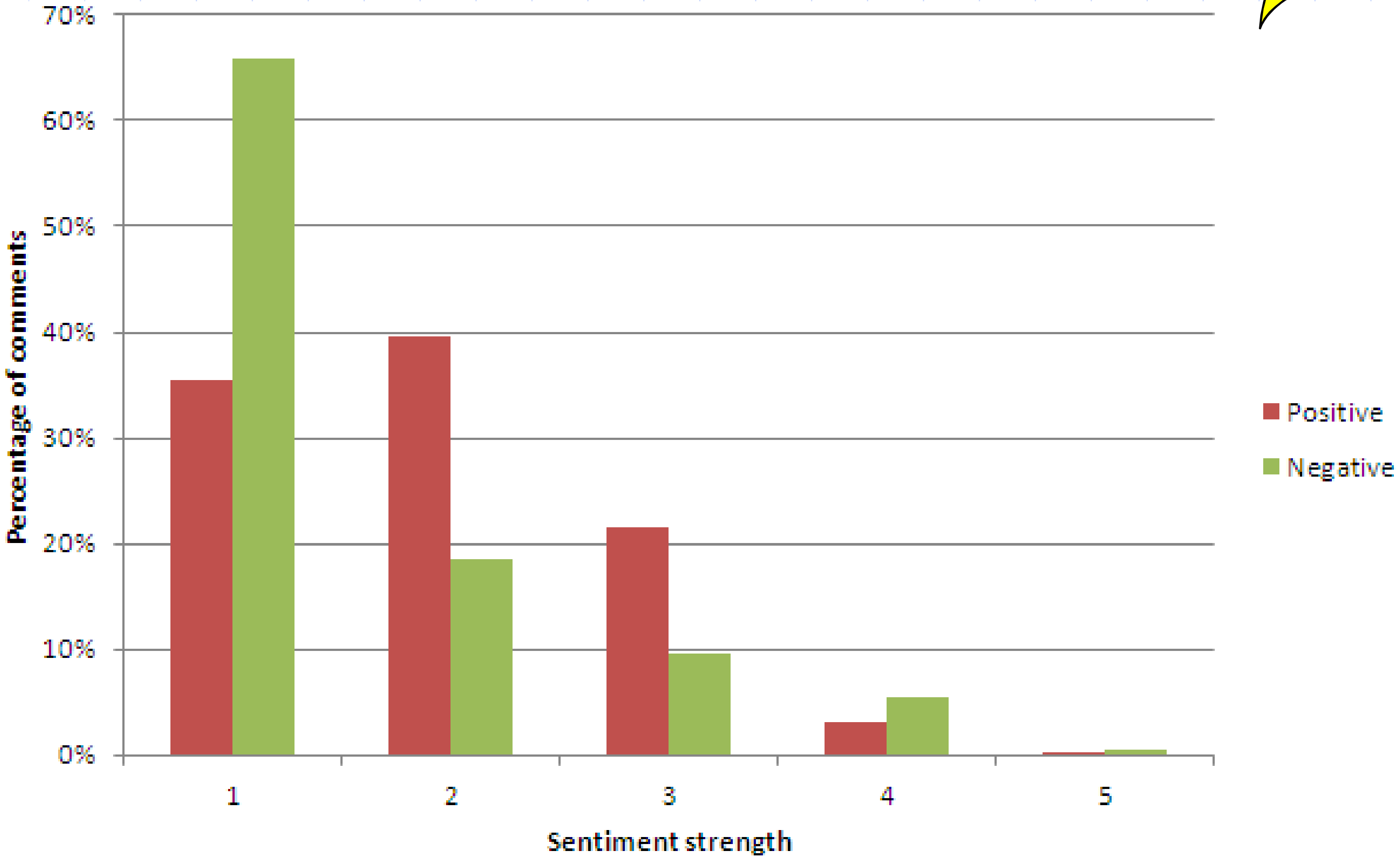
BB UR AMAZING :D

iTrolledABearOnce 19 minutes ago

Mike Thelwall Pardeep Sud Farida Vis (submitted)
Commenting on YouTube Videos: From Guatemalan Rock to El Big Bang

Sentiment in YouTube comments





Predominantly positive comments

Trends in YouTube comment sentiment

- +ve and –ve sentiment strengths negatively correlated for videos (Spearman's rho -0.213)
- # comments on a video correlates with –ve sentiment strength (Spearman's rho 0.242, p=0.000) and negatively correlates with +ve sentiment strength (Spearman's rho -0.113) – **negativity** drives commenting even though it is rare!
- Qualitative: Big debates over religion
  - No discussion about aging rock stars!

Conclusion

- Automatic classification of sentiment strength is possible for the social web – even unsupervised!
  - ...not good for longer, political messages?
- Hard to get accuracy much over 60%?
- Can identify trends through automatic analysis of sentiment in millions of social web messages

CYBEREMOTIONS

Bibliography

- Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12), 2544–2558.
- Thelwall, M., Buckley, K., & Paltoglou, G. (2011). Sentiment in Twitter events. *Journal of the American Society for Information Science and Technology,* 61(12), 2544–2558.