Semantic Sentiment Analysis of Twitter

Hassan Saif, Yulan He & Harith Alani

Knowledge Media Institute, The Open University, Milton Keynes, United Kingdom

The 11th International Semantic Web Conference ISWC 2012

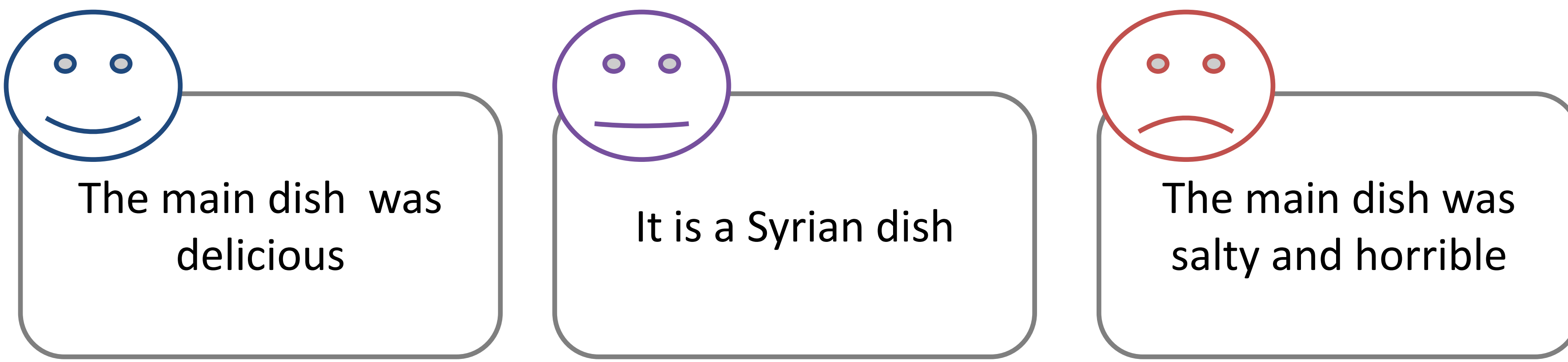
Boston- USA

Outline

- Background
- Twitter Sentiment Analysis
- Related Work
- Semantic Sentiment Analysis
- Evaluation
- Demo
- Conclusion
- Future Work

Sentiment Analysis

"Sentiment analysis is the task of identifying positive and negative opinions, emotions and evaluations in text"

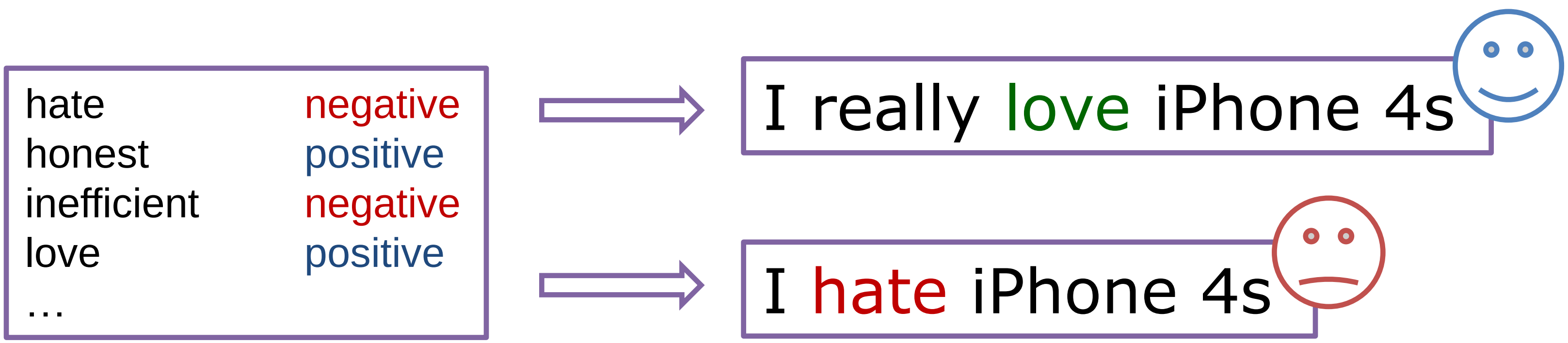


Opinion | Fact | Opinion

Sentiment Analysis

Lexical-Based Approach

- Building a better dictionary

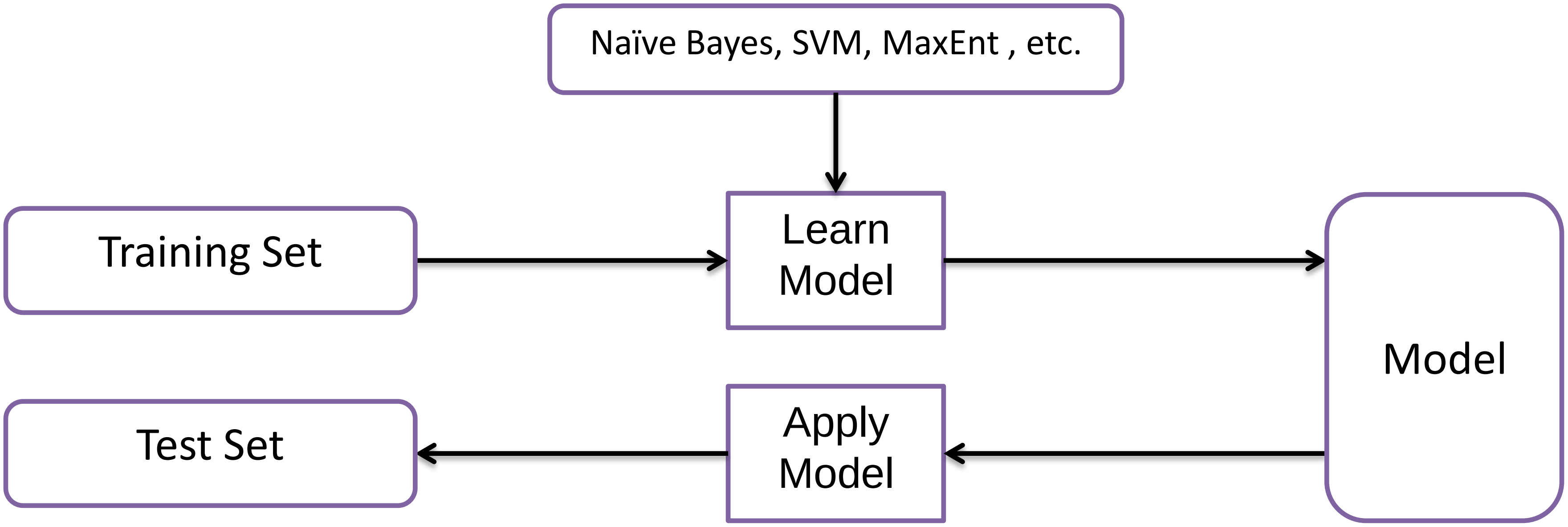


Sentiment Lexicon

Sentiment Analysis

Machine Learning Approach

- Finding the Right Feature



Twitter Sentiment Analysis

Challenges

- The short length of status update
- Language Variations
- Open Domain

Twitter Sentiment Analysis

Motivation

2010 U.S. Midterm Elections.
Conover et al. (2011)

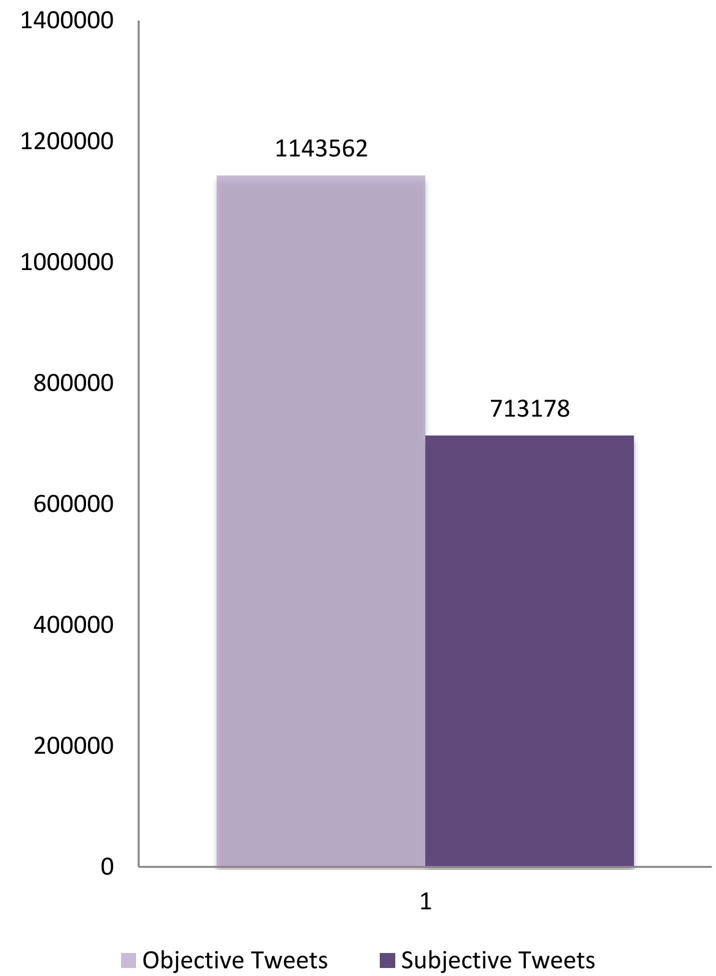
Stock Market Behavior
Bollen et al. (2010)

U.S. Presidential TV Debate in 2008

DIAKOPOULOS, N., AND SHAMMA, D. (2010)

UK General Elections, 2010

He & Saif (2012)



UK General Elections Corpus

Related Work

Twitter Sentiment Analysis

Related Work

- Distant Supervision
  - Supervised classifiers trained from noisy labels
  - Tweets messages are labeled using emoticons
  - Data filtering process

Go et al., (2009) - Barbosa and Fengl. (2010) – Pak and Paroubek (2010)

Twitter Sentiment Analysis

Related Work

- Followers Graph & Label Propagation
  - Twitter follower graph (users, tweets, unigrams and hashtags)
  - Start with small number of labeled tweets
  - Applied label propagation method throughout the graph.

Speriosu et al., (2009)

Twitter Sentiment Analysis

Related Work

- Feature Engineering
  - Unigrams, bigrams, POS
  - Microblogging features
    - Hashtags
    - Emoticons
    - Abbreviations & Intensifiers

Agsrwal et al., (2011) – Kouloumpis et al (2011)

Semantics

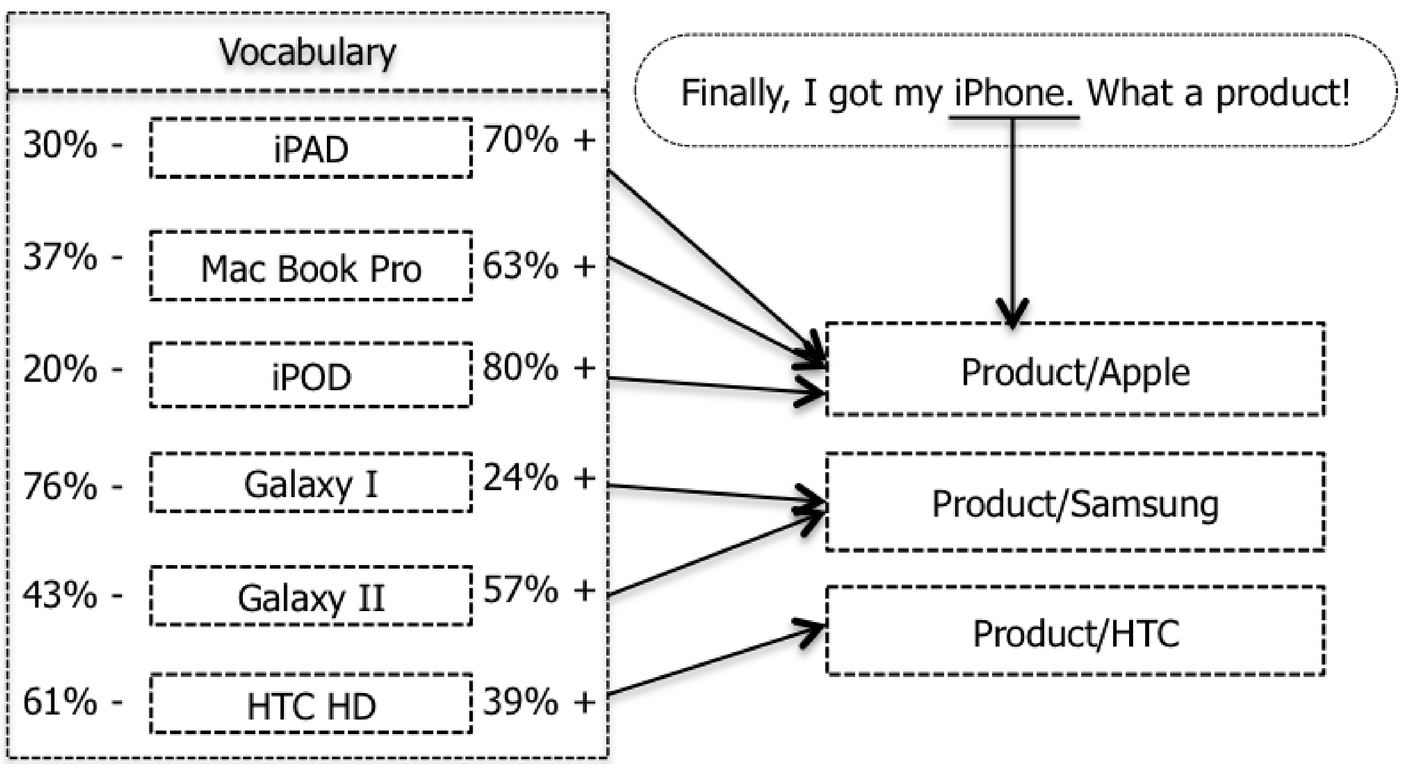
Semantic Sentiment Analysis

The Idea

Extract semantic concepts from tweets data and incorporate them into the supervised classifier training.

Semantic Sentiment Analysis

The Idea



Semantic Features Incorporation

(1) Shallow Semantic Methods (Replacement, Augmentation)

@Stace_meister Ya, I have **Rugby** in an hour

Sport

Sushi time for fabulous **Jesse**'s last day on dragons den

Person

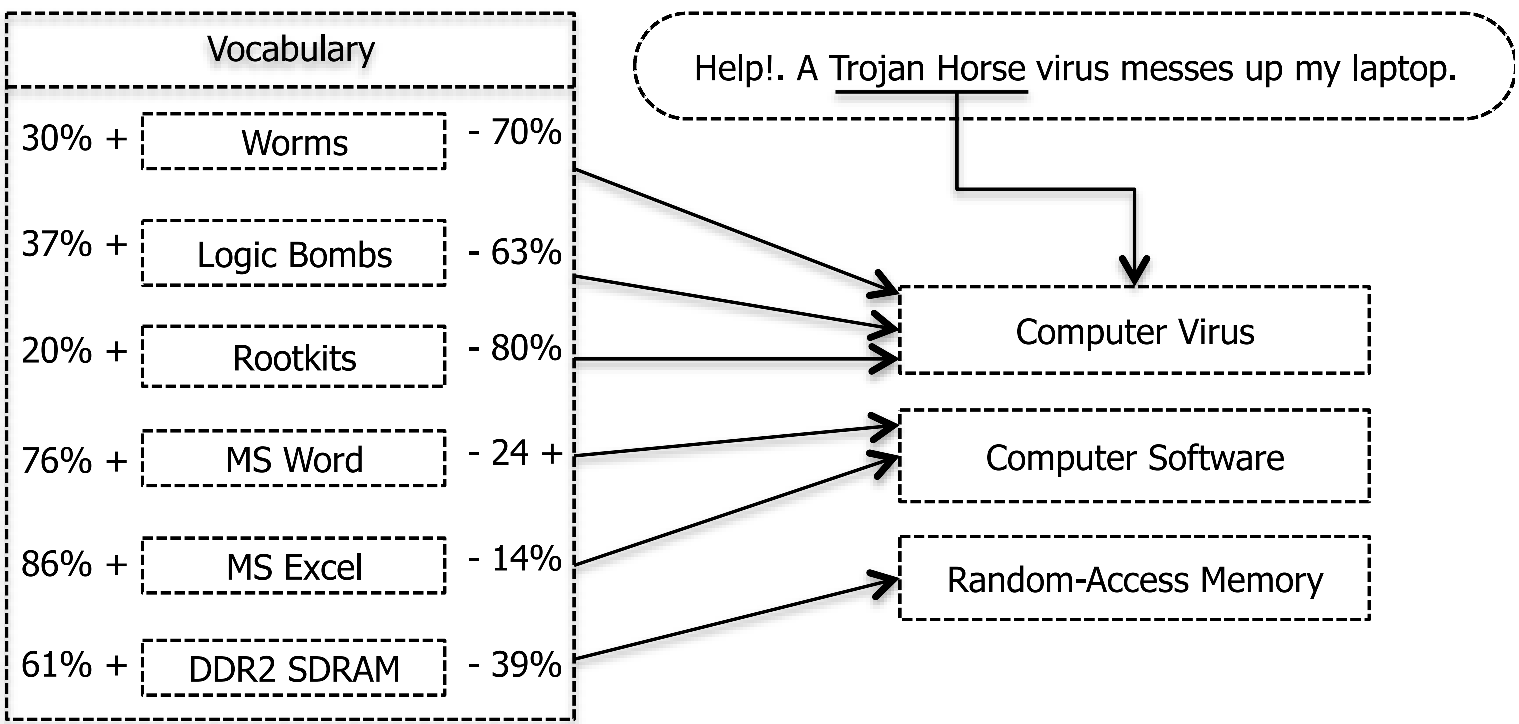
Dear **eBay**, if I win I owe you a total 580.63 bye paycheck

Company

Semantic Features Incorporation

Interpolation Method

$$P_s(w|c) = (1-\alpha)P_u(w|c) + \alpha \sum_j P(w|s_j)P(s_j|c)$$



Semantic Concept Extraction

- Using three different third-party tools: Zemanta, OpenCalais and AlchemyAPI
- AlchemyAPI extracted the most number of concepts and has the highest entity-concept mapping accuracy

Extraction Tool	No. of Concepts Extracted	Entity-Concept Mapping Accuracy (%)		
		Evaluator 1	Evaluator 2	Evaluator 3
AlchemyAPI	108	73.97	73.8	72.8
Zemanta	70	71	71.8	70.4
OpenCalais	65	68	69.1	68.7

Table 2. Evaluation results of AlchemyAPI, Zemanta and OpenCalais.

Evaluation

Datasets

Dataset	Type	No. of Tweets	Positive	Negative
Stanford Twitter Sentiment Corpus (STS)	Train	60K	30K	30K
	Test	1,000	470	530
Health Care Reform (HCR)	Train	839	234	421
	Test	839	163	536
Obama-McCain Debate (OMD)	n-fold cross validation	1,081	393	688

Evaluation

Baselines

I Like the new iPad

– Unigram Features:
[I, like, the, new, iPad]

– Part-of-Speech Features
[I(p), like(v), the(d), new(a), iPad(n)]

– Sentiment-Topic Features
[I, like, the, new, iPad] [(1_0), (1_1), (1_1), (1_0)]

Evaluation

Semantic Features Incorporation (F-measure)

Method	STS	HCR	OMD	Average
Semantic replacement	74.10	61.35	71.25	68.90
Semantic augmentation	77.65	63.65	72.70	71.33
Semantic interpolation	**83.90**	**66.10**	**77.85**	**75.95**

Average sentiment classification accuracy (%) using different methods for incorporating the semantic features. Accuracy here is the average harmonic mean (F measure) obtained from identifying positive and negative sentiment.

Evaluation

Cross Comparison (F-Measure)

Dataset	Feature	Positive Sentiment			Negative Sentiment			Average		
		P	R	F1	P	R	F1	P	R	F1
STS	Unigrams	82.20	75.20	78.50	79.30	85.30	82.20	80.75	80.25	80.35
	POS	83.70	75.00	79.10	79.50	86.90	83.00	81.60	80.95	81.05
	Sentiment-Topic	80.70	82.20	81.40	83.70	82.30	83.00	82.20	82.25	82.20
	Semantics	85.80	79.40	82.50	82.70	88.20	85.30	**84.25**	**83.80**	**83.90**
HCR	Unigrams	39.00	36.10	37.50	81.00	82.80	81.90	60.00	59.45	59.70
	POS	56.20	22.00	31.70	80.00	94.70	86.70	68.10	58.35	59.20
	Sentiment-Topic	53.80	47.20	50.30	84.50	87.60	86.00	**69.15**	**67.40**	**68.15**
	Semantics	53.60	40.40	46.10	83.10	89.30	86.10	68.35	64.85	66.10
OMD	Unigrams	64.20	70.90	67.10	83.30	78.60	80.80	73.75	74.75	73.95
	POS	69.50	68.30	68.70	83.10	83.90	83.40	76.30	76.10	76.05
	Sentiment-Topic	68.20	75.60	71.70	87.10	82.40	84.70	77.65	**79.00**	**78.20**
	Semantics	75.00	66.60	70.30	82.90	88.10	85.40	**78.95**	77.35	77.85

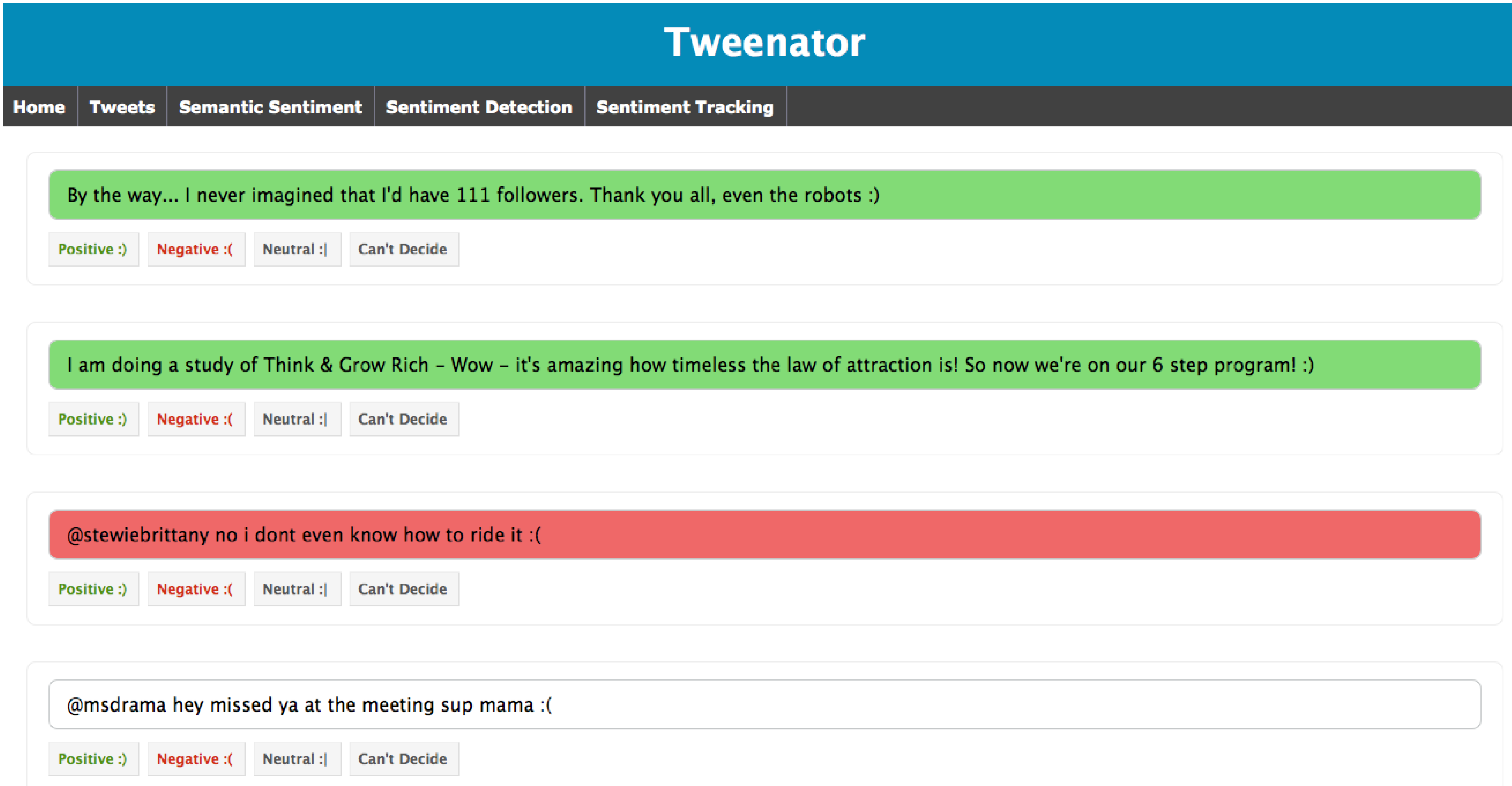
Evaluation

Cross Comparison (F-Measure)

Features	Positive Sentiment			Negative Sentiment			Average		
	P	R	F1	P	R	F1	P	R	F1
Unigrams	61.80	60.73	61.03	81.20	82.23	81.63	71.50	71.48	71.33
POS	69.80	55.10	59.83	80.87	88.50	84.37	75.53	72.23	72.48
Sentiment-Topic	67.57	**68.33**	**67.80**	**85.10**	84.10	84.57	77.02	**76.73**	**76.75**
Semantics	**71.47**	62.13	66.30	82.90	**88.53**	**85.60**	**77.18**	75.33	75.95

Table 9. Averages of Precision, Recall, and F measures across all three datasets.

Tweenator



www.tweenator.com

Conclusion

- Twitter Sentiment Analysis is very challenging problem.
- We proposed using semantic features for Twitter SA using three different methods: replacement, augmentation and interpolation.
- We found that the interpolation method outperforms the other two methods.
- We compared our features with three different baselines and showed that semantic features on average are more precise amongst them.
- There is no winning approach. The accuracy of classifying with some feature selections can be sensitive to the size of the datasets and their topical-focus.

Future Work

- Extracting Semantic Entities and Concepts
  - Explore more fine-grained approach for the entity extraction and the entity-concept mapping
- Selective Interpolation Method
  - Interpolate semantic concepts based on their contribution to the classification performance.
- 3-way Sentiment Analysis
  - Propose a hyper classifier that is able to work with Objective and Subjective tweets.

References

[1] AGARWAL, A., XIE, B., VOVSHA, I., RAMBOW, O., AND PASSONNEAU, R. Sentiment analysis of twitter data. In Proceedings of the ACL 2011 Workshop on Languages in Social Media (2011), pp. 30–38.

[2] BARBOSA, L., AND FENG, J. Robust sentiment detection on twitter from biased and noisy data. In Proceedings of COLING (2010), pp. 36–44.

[3] BIFET, A., AND FRANK, E. Sentiment knowledge discovery in twitter streaming data. In Discovery Science (2010), Springer, pp. 1–15.

[4] GO, A., BHAYANI, R., AND HUANG, L. Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford (2009).

[5] KOULOUMPIS, E., WILSON, T., AND MOORE, J. Twitter sentiment analysis: The good the bad and the omg! In Proceedings of the ICWSM (2011).

[5]LIN, C., AND HE, Y. Joint sentiment/topic model for sentiment analysis. In Proceeding of the 18th ACM conference on Information and knowledge management (2009), ACM, pp. 375–384.

[6] PAK, A., AND PAROUBEK, P. Twitter as a corpus for sentiment analysis and opinion mining. Proceedings of LREC 2010 (2010).

[7]SAIF, H., HE, Y., AND ALANI, H. Semantic Smoothing for Twitter Sentiment Analysis. In Proceeding of the 10th International Semantic Web Conference (ISWC) (2011).

[8] Saif, H., He, Y. and Alani, H. (2012) Alleviating Data Sparsity for Twitter Sentiment Analysis, Workshop: The 2nd Workshop on Making Sense of Microposts (#MSM2012): Big things come in small packages at World Wide Web (WWW) 2012, Lyon, France

Thank You

Email: hassan.saif@open.ac.uk
Twitter: hrsaif
Website: tweenator.com