

User Level Sentiment Analysis Incorporating Social Networks

Chenhao Tan

Department of Computer Science
Cornell University

Joint work with :
Lillian Lee, Jie Tang, Long Jiang, Ming Zhou and Ping Li

May 18, 2011

Outline

- 1 Motivation
- 2 Problem Setting in Twitter
- 3 Data Collection
- 4 Observation
- 5 Model
- 6 Approach
- 7 Experiment
- 8 Conclusion

Motivation

- User-level sentiment analysis
- Network information
 - Accessibility
 - Homophily or Attention

Twitter as the basis

- Text information: Tweets
- Network information
 - Follow Network
 - @ Network
 - ◊ directed
 - ◊ mutual

Semi-supervised Learning in Twitter

- Hard to get full labels
- Given a graph and labels of some nodes in the graph, try to classify the other users in the graph

- Traditional Annotation by Tweets

- **Failed** Traditional Annotation by Tweets

- **Failed** Traditional Annotation by Tweets
- User Biographical Information

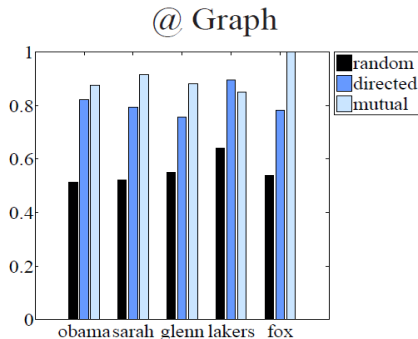
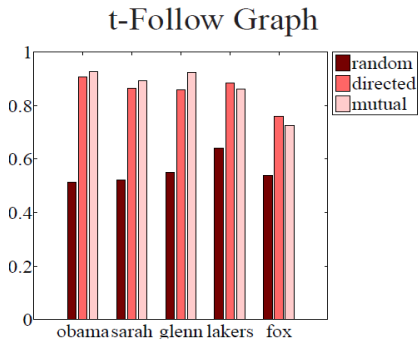
Final Data Set

- 1, 414, 340 users
- 1, 414, 211 user profiles
- 480, 435, 500 tweets
- 274, 644, 047 t-follow edges
- 58, 387, 964 @-edges

Topic	#labeled users	#t-follow edges (mutual)	#@ edges (mutual)	#users in total	total # of on-topic tweets
Obama	889	7838(2949)	2358(302)	307985	4873711
Sarah Palin	310	1003(264)	449(60)	121910	972537
Glenn Beck	313	486(159)	148(17)	95847	687913
Lakers	640	2297(353)	1167(127)	76926	301558
Fox News	231	130(32)	37(5)	114530	1231519

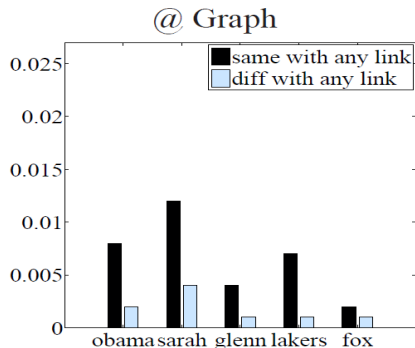
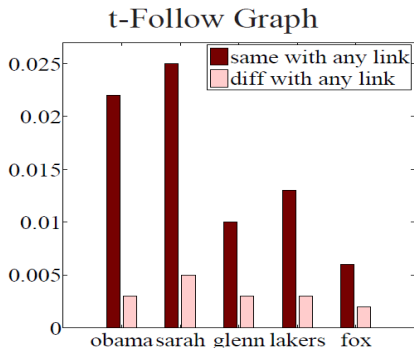
Sharing Label conditioned on being connected

- Probability that two users have the same label, conditioned on whether or not they are connected



Connectedness conditioned on labels

- Probability that two users are connected, conditioned on whether or not they have the same label



Model Framework

- User-Tweet Factor

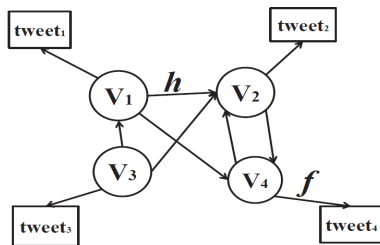
$$f_{k,l}(y_i, y_t) = \begin{cases} \frac{w_{\text{labeled}}}{|\text{tweet}_{v_i}|} & y_i = k, y_t = l \\ \frac{w_{\text{unlabeled}}}{|\text{tweet}_{v_i}|} & y_i = k, y_t = l \\ 0 & \text{otherwise} \end{cases}$$

- User-User Factor

$$h_{k,l}(y_i, y_j) = \begin{cases} \frac{w_{\text{relation}}}{|\text{Neighbors}_{v_i}|} & y_i = k, y_j = l \\ 0 & \end{cases}$$

- Objective Function

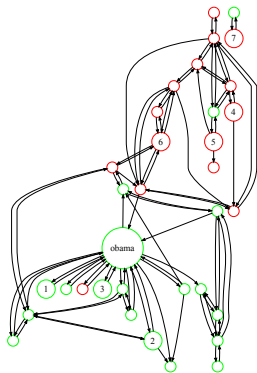
$$\begin{aligned} \log P(\mathbf{Y}) = & \left(\sum_{v_i \in V} \left[\sum_{t \in \text{tweet}_{v_i}, k, l} \mu_{k,l} f_{k,l}(y_{v_i}, y_t) + \right. \right. \\ & \left. \left. \sum_{v_j \in \text{Neighbors}_{v_i}, k, l} \lambda_{k,l} h_{k,l}(y_{v_i}, y_{v_j}) \right] \right) \\ & - \log Z \end{aligned}$$



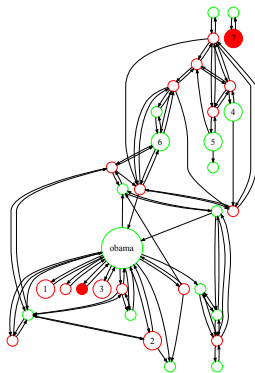
- Parameter Estimation
 - Direct estimation from simple statistics
 - SampleRank
- Inference
 - loopy belief propagation

- Training set: 50 positive users and 50 negative users
the others for testing
- Labels of Tweets
 - SpecificSVM
- Labels of Users
 - Majority Vote
 - HGM-NoLearning
 - HGM-Learning

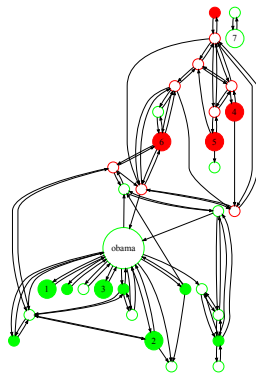
Case Study



(a) Ground Truth



(b) Text-Only Approach

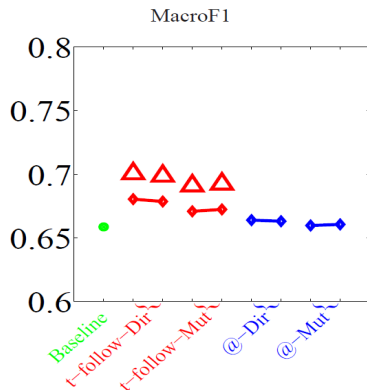
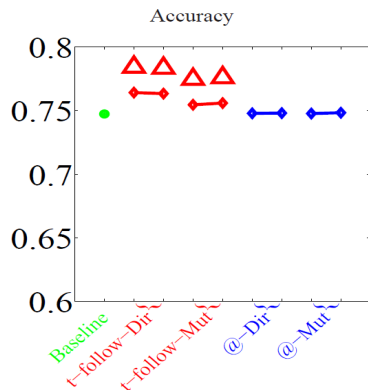


(c) Our algorithm

Sample tweets of users classified correctly only with network information

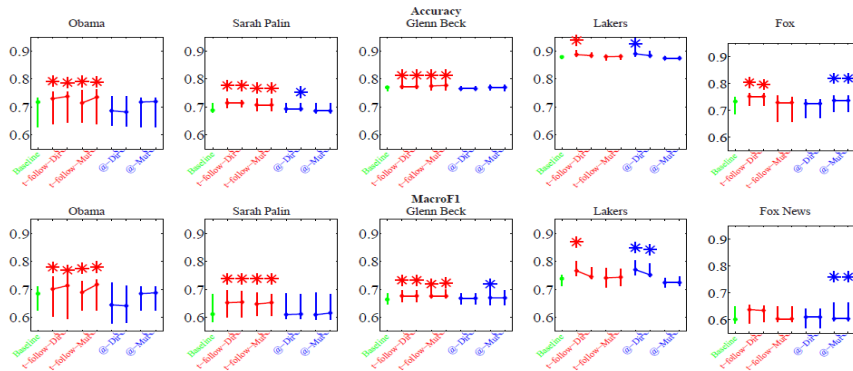
User ID	SVM	HGM	True	Tweet
1	NEG	POS	POS	Obama is making the repubs look silly and petty. #hrc
2	NEG	POS	POS	Is happy Obama is President
				Obama collectable http://tinyurl.com/c5u7jf
3	NEG	POS	POS	I am praying that the government is able to get health care reformed this year! President Obama seems like the ONE to get it worked out!!
				Watching House on TV. I will be turning to watch Rachel M. next. I am hoping Pres. Obama gets his budget passed. Especially Health Care!
4	POS	NEG	NEG	RT @TeaPartyProtest Only thing we have 2 fear is Obama himself & Pelosi & Cong & liberal news & Dems &... http://ow.ly/15M9Xv
				RT @GlennBeckClips: Barack Obama can no more disown ACORN than he could disown his own grandmother. #TCOT
5	POS	NEG	NEG	RT @JosephAGallant Twitlonger: Suppose I wanted to Immigrant to Mexico? A Letter to President Obama.. http://tl.gd/1kr5rh
				George Bush was and acted like a war time President. Obama is on a four year power grab and photo op. #tcot
6	POS	NEG	NEG	ObamaCare forces Americans to buy or face a fine! It is UNCONSTITUTIONAL to force us to buy obamacare. Marxist Govt. taking our Freedoms!
				Look up Chicago Climate Exchange,an organization formed years ago by Obama & his Marxist-Commie Cronies to form a profit off cap & trade.

Overall Performance



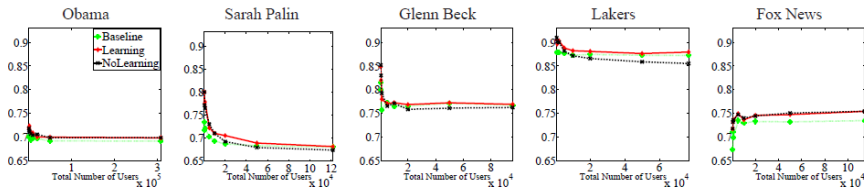
- Beat Baseline!
- *Follow* better than *@*
- *Directed* better than *Undirected*
- NoLearning same with Learning

Performance Per Topic



- Sparseness of graph
- Size of graph or #Tweets per user
- SVM Classifier Performance

Adding More Unlabeled Data



- Learning better than NoLearning

Conclusion

- Empirical analyses on the correlation of networks and sentiment
- Propose a heterogeneous graphical model
- Validate the effectiveness of incorporating network information

- More data sets
- Better models and semi-supervised learning algorithms
- Find the helpful parts of networks
- Build a theory of why and how users correlate on different topics in different kinds of networks

The End

Thank you!
Questions?