

Analyzing Social Media Content for Security Informatics

Rich Colbaugh

Kristin Glass

Sandia National Laboratories

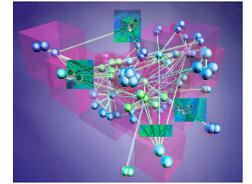
August 2013

This work was funded in part by the U.S. Food and Drug Administration through a contract with the U.S. Department of Energy/Sandia National Laboratories (funding document 224109011).

The information in this presentation is not a formal dissemination of information by the FDA and does not represent agency position or policy. The contents are the responsibility of the authors alone.



Introduction



Background

- Web-users regularly offer their views and opinions concerning security-relevant topics, for instance political protests or potential epidemics.
- There is considerable interest in leveraging this information to support various objectives (e.g. threat warning, disease outbreak surveillance).
- Moreover, there is evidence for the feasibility of this general notion:
 - *non-security* – forecasting [Colbaugh/Glass 2010, Asur/Huberman 2010, Berger et al. 2010, Bollen et al. 2011, Amodea et al. 2011]; monitoring/surveillance [Melville et al. 2009, Maniu et al. 2011, Ayers et al. 2011, Colbaugh/Glass 2012, Bhattacharya et al. 2012, Kosinski et al. 2013];
 - *security* – forecasting [Colbaugh/Glass 2012a, 2012b, Sadilek et al. 2012]; monitoring/surveillance [Culotta 2010, Lampos et al. 2010, Glass/Colbaugh 2011].



Introduction



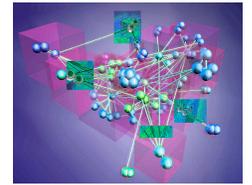
Sentiment/emotion analysis

- Public sentiment and emotion regarding issues and events, and the way it is distributed, is of particular interest in many settings, such as when predicting the threat level posed by a contentious situation or likelihood that a new vaccine will be adopted.
- However, accurately estimating sentiment/emotion expressed in social media is challenging for many reasons, including
 - *volume* – nearly 1B users generate content each day;
 - *style* – content is typically expressed using informal/imprecise language;
 - *labels* – machine learning methods are promising but acquiring training data is expensive/time consuming.





Introduction



Objective

Develop accurate, scalable, flexible, and easy-to-implement approach to estimating the sentiment and/or emotion of social media content.

Outline

- Problem formulation:
tasks of interest, content analysis via machine learning.
- Proposed approach:
Algorithm SEE (sentiment/emotion estimation), empirical evaluation.
- Case study: Israel/Palestine conflict.
- Case study: H1N1 (flu) epidemic.





Problem Formulation

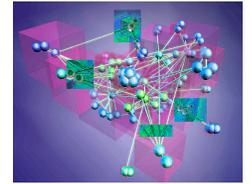


Setup

- Tasks of interest:
 - *document classification* – accurately estimate sentiment or emotion polarity of a particular document;
 - *corpus classification* – accurately estimate sentiment or emotion distribution for a collection of documents.
- Input: small, generic lexicon of sentiment-laden/emotion-laden words and corpus of unlabeled documents.
- Key considerations:
 - *generality/flexibility* (e.g. method should be domain independent and able to adapt to novel communication modes (e.g. slang));
 - *convenience/expense* (e.g. method should be implementable with *no* labeled training documents).

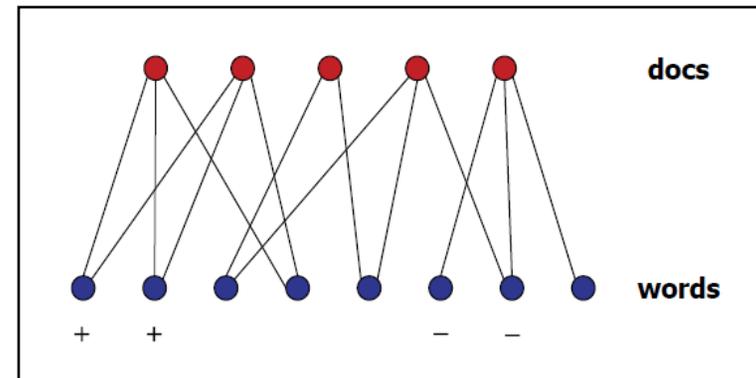


Problem Formulation



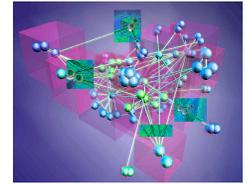
Sentiment/emotion via machine learning

- Problem: given a 1.) corpus of n documents of unknown sentiment or emotion polarity composed of words from some vocabulary V , and 2.) modest lexicon $V_1 \subseteq V$ of words of known sentiment/emotion polarity, estimate the polarity of all the documents.
- Approach: leverage information in unlabeled documents by
 - modeling the data as a bipartite graph G_b (words V_1 are labeled);
 - assuming that, in G_b , positive/negative documents tend to be connected to positive/negative words;
 - learning document polarity via graph transduction.





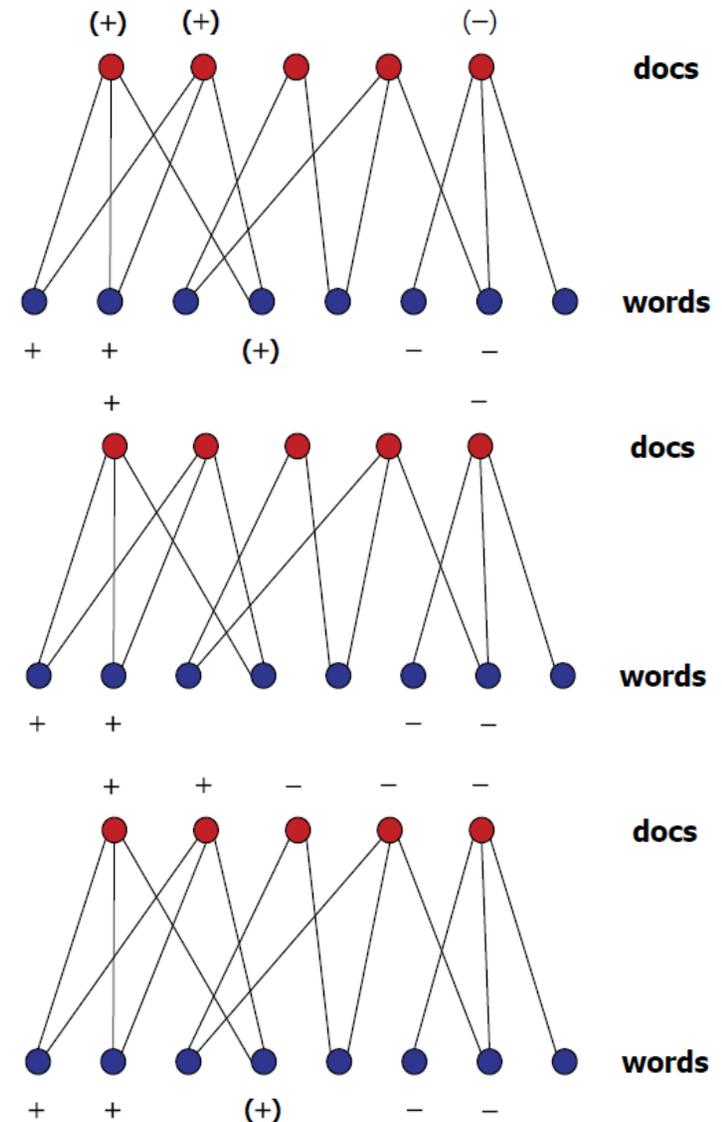
Proposed Approach



Algorithm SEE

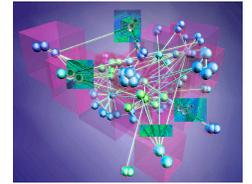
Basic idea

1. Construct intermediate classifier C: doc \rightarrow class using semi-supervised bipartite graph-based learning [Colbaugh/Glass, EISIC 2011].
2. Employ C to estimate polarity of all documents; label those documents about which C is “confident”.
3. Obtain final document polarity estimates via graph transduction on partially-labeled bipartite graph data model.



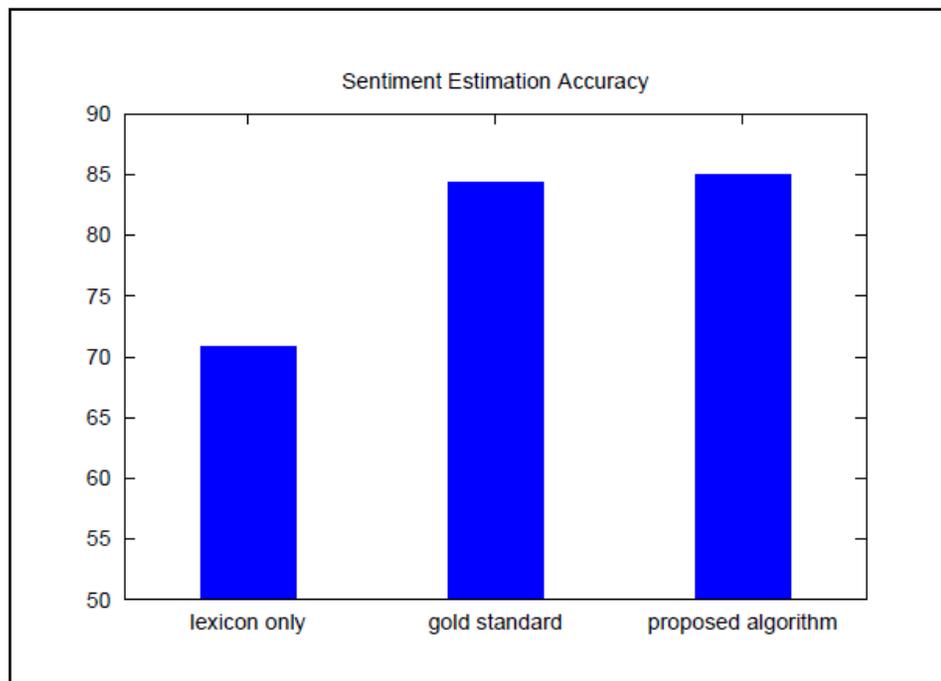


Proposed Approach



Empirical evaluation

Sentiment of online product reviews (“gold standard” is SCL algorithm of [Blitzer et al. 2007] trained on 1600 documents).



Sentiment Estimation

<u>method</u>	<u>accuracy</u>
lexicon-only	70.8%
gold standard	84.4%
Algorithm SEE	85.0%

Sentiment Proportion

<u>method</u>	<u>accuracy</u>
Algorithm SEE	95.3%



Case Study: Suicide Bombing



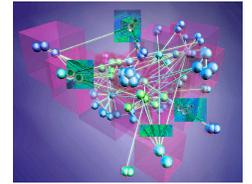
Introduction

- Recent research suggests characteristics of online discourse about contentious issues and events can be predictive of real-world behavior (e.g. protests and cyber attacks) [Colbaugh/Glass 2012].
- We explore this possibility by estimating, for the period 1997-2006, 1.) regional opinion regarding Palestinian suicide bombing against Israel, and 2.) suicide bombing attacks against Israel by Palestinian groups (Fatah, Hamas, PFLP, PIJ). Opinion is estimated by applying Algorithm SEE to online Arabic-language content (e.g. blog posts, editorials).



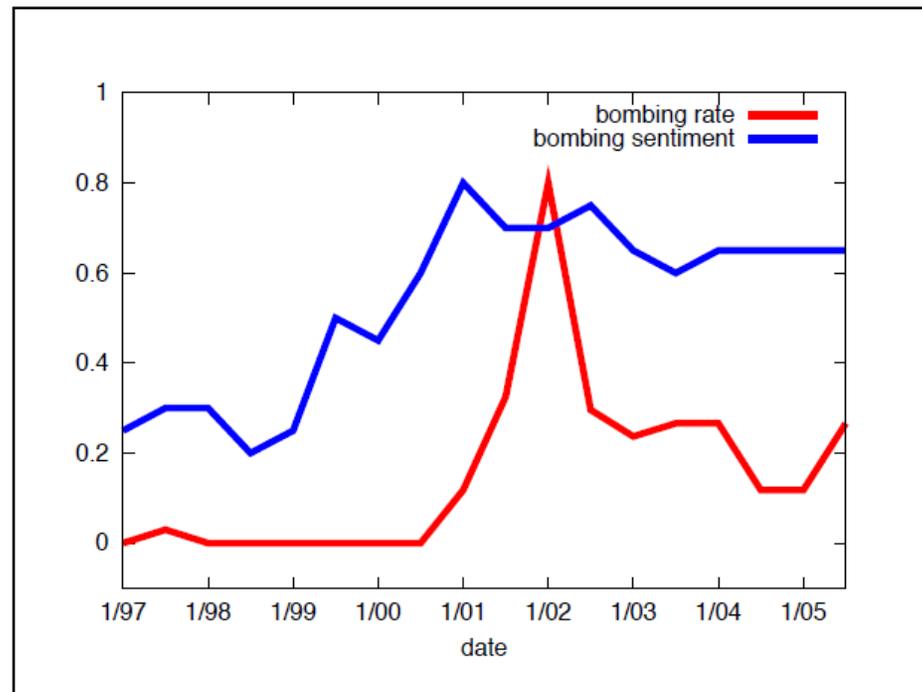


Case Study: Suicide Bombing



Online sentiment and suicide bombings

Regional sentiment about Palestinian suicide bombing (blue, normalized) and bombing frequency (red, normalized) are correlated, with sentiment leading bombing events by 12 months ($R = 0.8$, $p < 0.0001$).



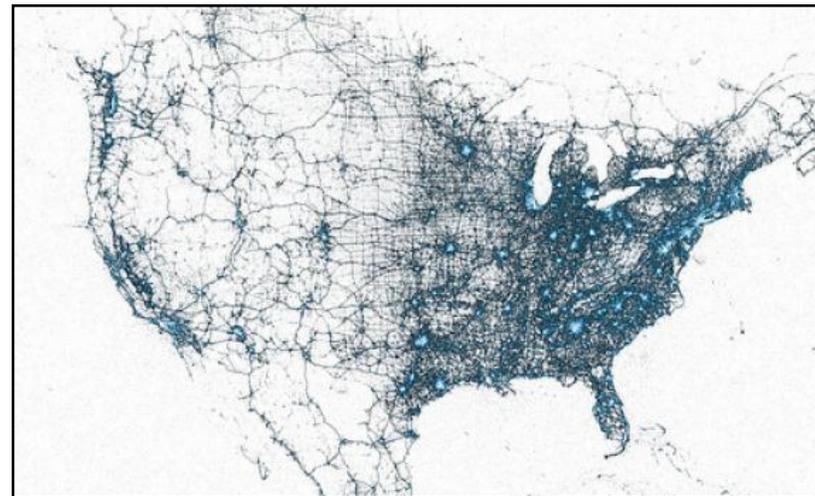


Case Study: H1N1 Epidemic



Introduction

- Public sentiment regarding vaccination can strongly affect vaccination rates [Signorini et al. 2011, Bhattacharya et al. 2012], and therefore the risk of disease outbreak and overall security of a population.
- We estimate public sentiment concerning vaccination for H1N1 flu during the second half of 2009 by applying Algorithm SEE to a dataset of ~500M Twitter posts and the associated “@-network” of user interactions.





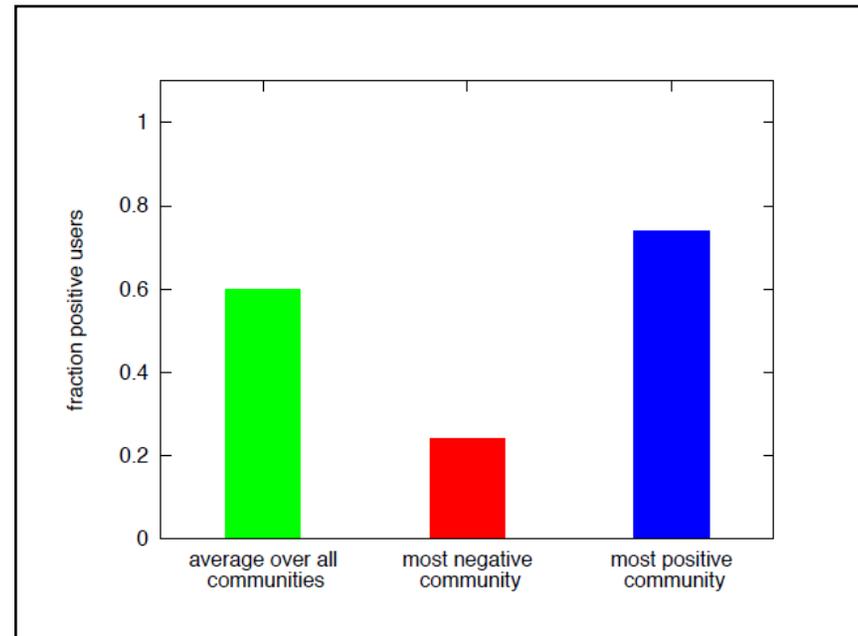
Case Study: H1N1 Epidemic



Vaccine sentiment distribution

Sentiment analysis reveals that public opinion regarding the H1N1 vaccine, while positive, is distributed heterogeneously over the communities of the Twitter graph, and that some communities have quite negative sentiment.

Sample results





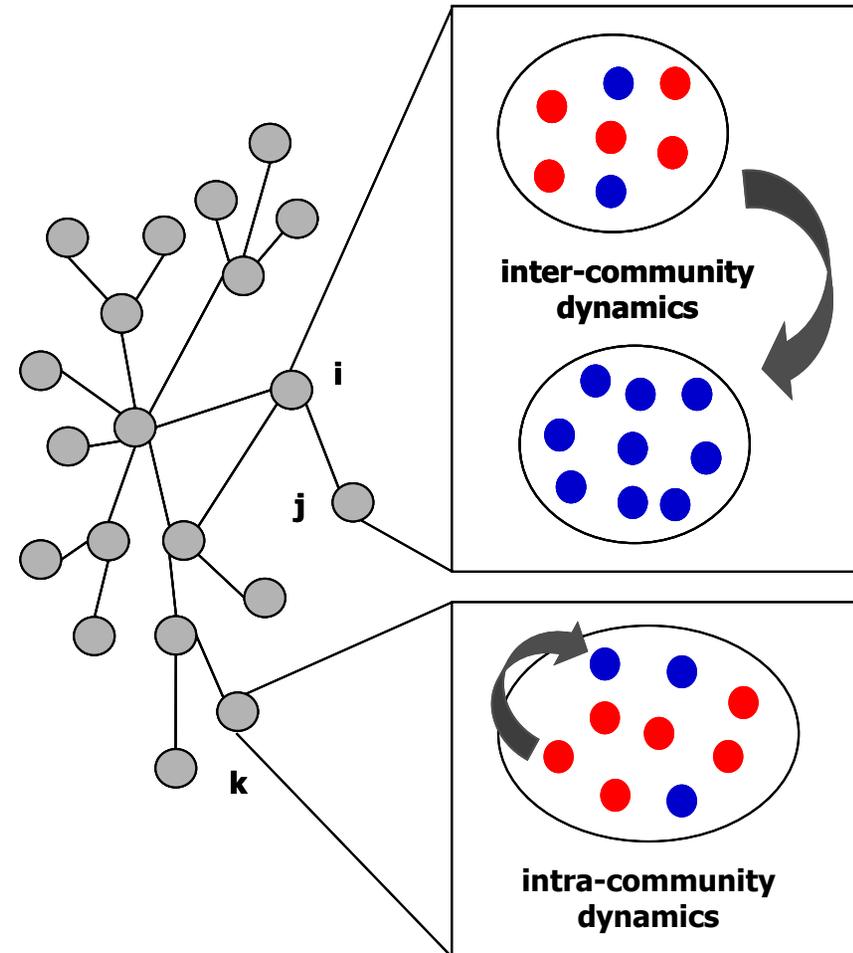
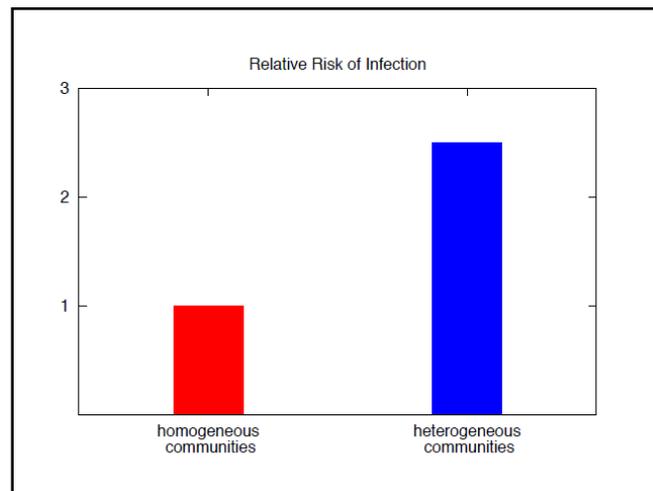
Case Study: H1N1 Epidemic



Implication of heterogeneous sentiment distribution

Simulation study

We explore the implications of a heterogeneous distribution of sentiment over a population's social network via simulation using a social dynamics model.





Appendix



Algorithm SEE: some details

1. Construct intermediate classifier C: $\text{polar} = \text{sign}(\mathbf{c}^T \mathbf{x})$, where $\mathbf{x} \in \mathfrak{R}^{|\mathcal{V}|}$ is a bag-of-words document model and $\mathbf{c} \in \mathfrak{R}^{|\mathcal{V}|}$ is learned via

$$\min_{\mathbf{c}_{\text{aug}}} \mathbf{c}_{\text{aug}}^T \mathbf{L}_n^k \mathbf{c}_{\text{aug}} + \beta \sum_{i=1}^{|\mathcal{V}_1|} (c_i - w_i)^2$$

where $\mathbf{c}_{\text{aug}} = [\mathbf{d}_{\text{est}}^T \ \mathbf{c}^T]^T$, L_n is the normalized Laplacian for G_b , and $\mathbf{w} \in \mathfrak{R}^{|\mathcal{V}|}$ encodes the lexicon \mathcal{V}_1 of words of known sentiment/emotion.

2. Assign preliminary labels to the (n_i) documents with large magnitude polarity estimates: if $d_{\text{est},i} > d_{\text{thresh}}$ ($d_{\text{est},i} < -d_{\text{thresh}}$) then $d_i = 1$ ($d_i = -1$), where $\mathbf{d} \in \mathfrak{R}^{n_i}$.



Appendix



Algorithm SEE: some details (cont' d)

3. Obtain the final document polarity estimates via graph transduction on the new version of G_b (possessing the n_1 preliminary document labels). Here graph transduction is performed by solving the following optimization problem:

$$\min_{c_{\text{aug}}} c_{\text{aug}}^T L_n^k c_{\text{aug}} + \beta_1 \sum_{i=1}^{n_1} (d_{\text{est},i} - d_i)^2 + \beta_2 \sum_{i=1}^{|V_1|} (c_i - w_i)^2$$

and \mathbf{d}_{est} is the final estimate for the (sentiment/emotion) polarity of all documents.