SAND2013-6887C

Analyzing Social Media Content for Security Informatics

Rich Colbaugh **Kristin Glass**

Sandia National Laboratories

March 2013

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]; "predicting the present" [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]; "predicting the present" [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 distribute, 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* – learning methods are promising but acquiring training data is expensive and time consuming.

Introduction

Objective

Develop accurate, flexible, scalable, 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

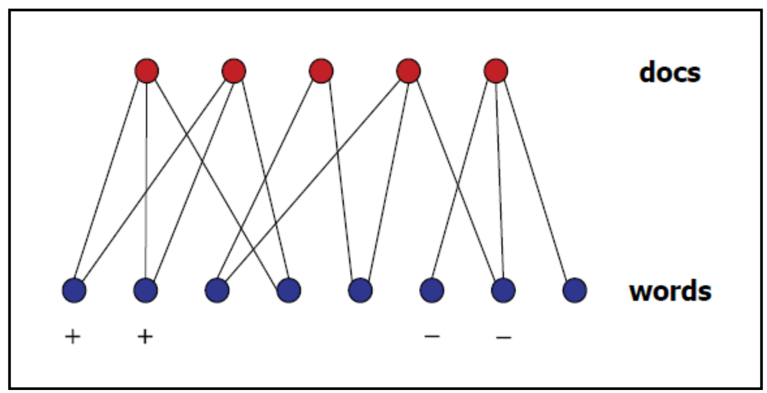
Setup

- Tasks of interest:
  - document classification – accurately estimate sentiment or emotion polarity of a particular document;
  - corpus classification – accurately estimate sentiment or emotion polarity of 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 example documents).

Problem Formulation

Sentiment/emotion via machine learning

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

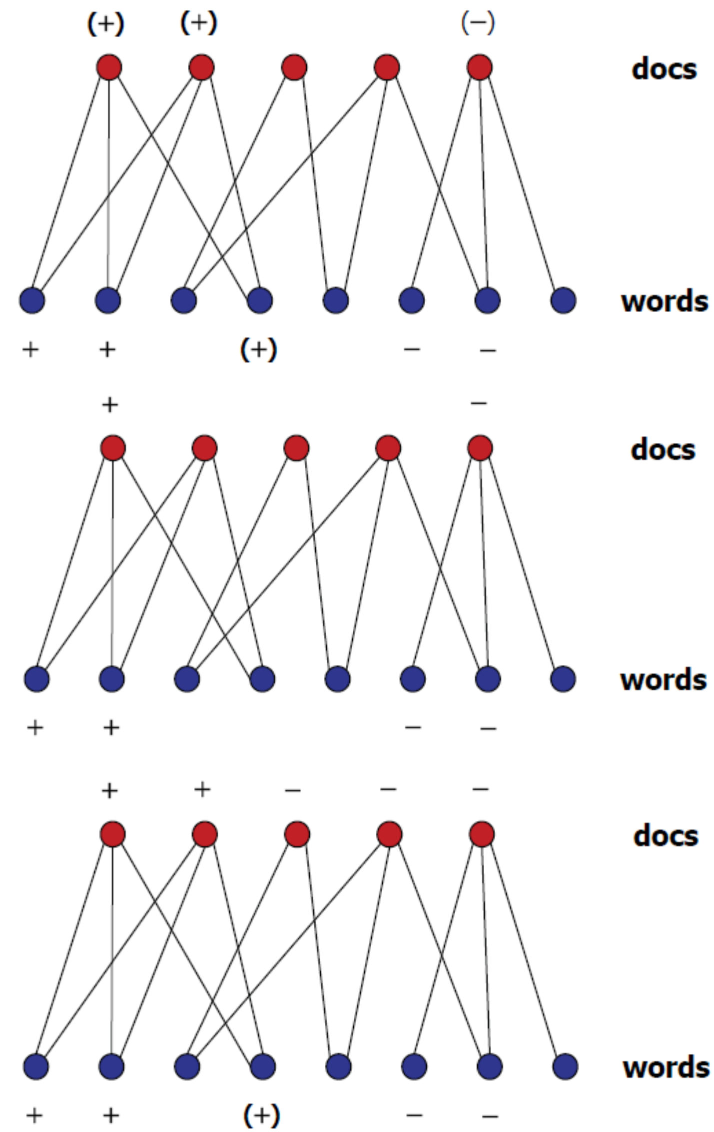


Proposed Approach

Algorithm SEE

Basic idea

1. Construct intermediate classifier C: doc $\rightarrow$ class using semi-supervised bipartite graph-based learning [Glass/Colbaugh 2012].
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 model.



Proposed Approach

Algorithm SEE (cont'd)

Mathematical details

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

$$\boxed{\min_{c_{aug}} \; c_{aug}^T L_n^k c_{aug} + \beta \sum_{i=1}^{|V_l|} (c_i - w_i)^2}$$

where $\mathbf{c}_{aug} = [\mathbf{d}_{est}^T \;\; \mathbf{c}^T]^T$ and L_n is the normalized Laplacian for G_b.

2. Assign preliminary labels to the (n_l) documents with large magnitude polarity estimates: if $d_{est,i} > d_{thresh}$ ($d_{est,i} < -d_{thresh}$) then $d_i = 1$ ($d_i = -1$), where $d \in \Re^{nl}$ and the document indices are reordered so that the first n_l are the ones assigned these preliminary labels.

Proposed Approach

Algorithm SEE (cont'd)

Mathematical details (cont'd)

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

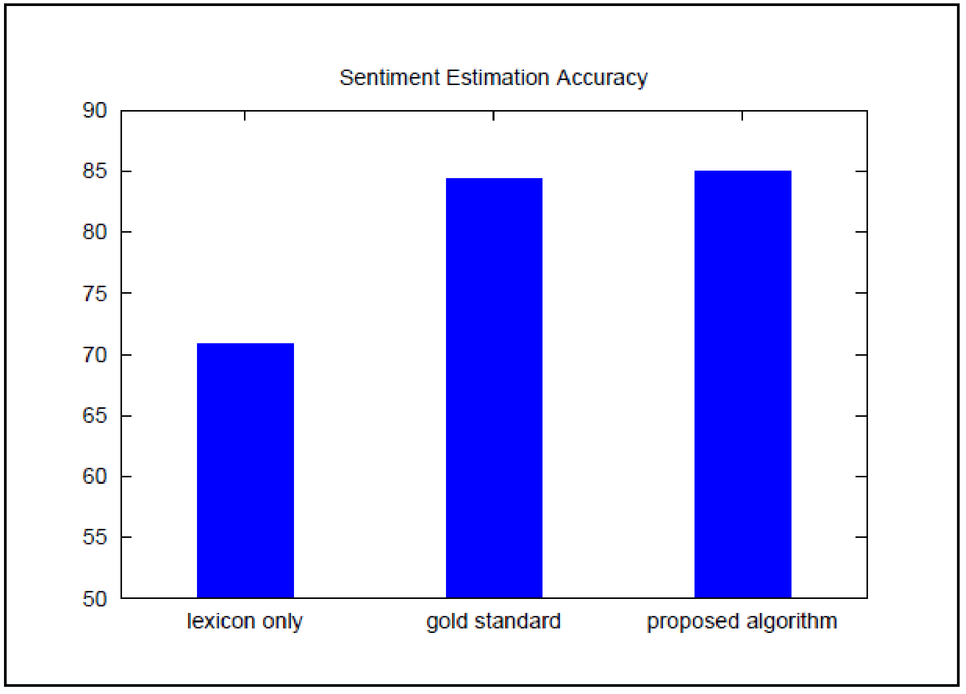
$$\boxed{\min_{c_{aug}} \; c_{aug}^T L_n^k c_{aug} + \beta_1 \sum_{i=1}^{n_l} (d_{est,i} - d_i)^2 + \beta_2 \sum_{i=1}^{|V_l|} (c_i - w_i)^2}$$

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

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

method	**accuracy**
lexicon-only	**70.8%**
gold standard	**84.4%**
Algorithm SEE	**85.0%**

Sentiment Proportion

method	**accuracy**
Algorithm SEE	**95.3%**

Case Study: Israel/Palestine Conflict

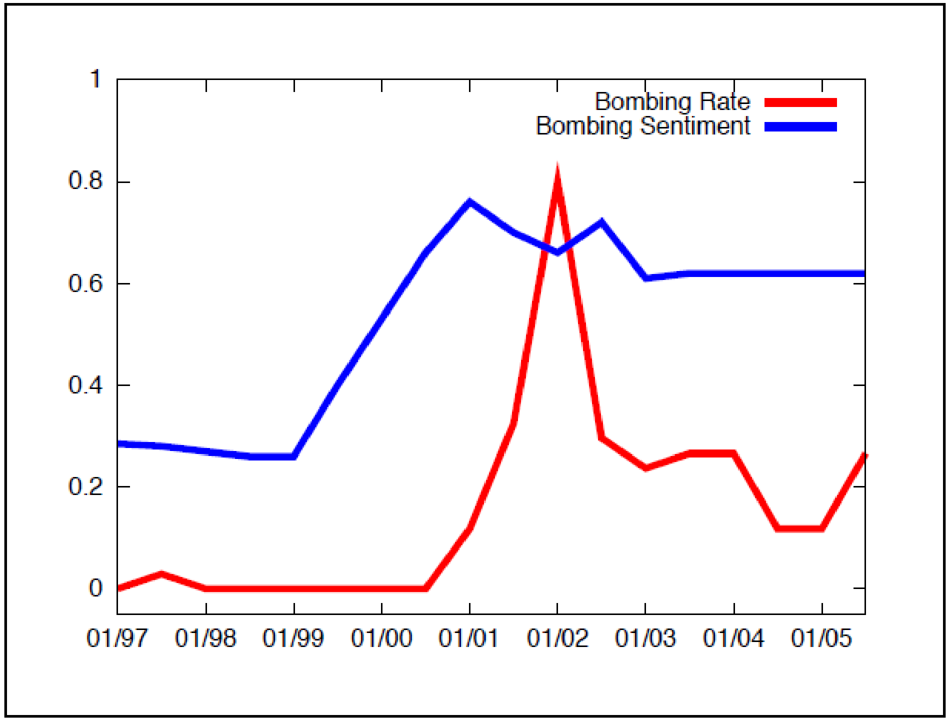
Introduction

- 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 Arabic-language content (e.g. blog posts, editorials).

Case Study: Israel/Palestine Conflict

Online sentiment and suicide bombings

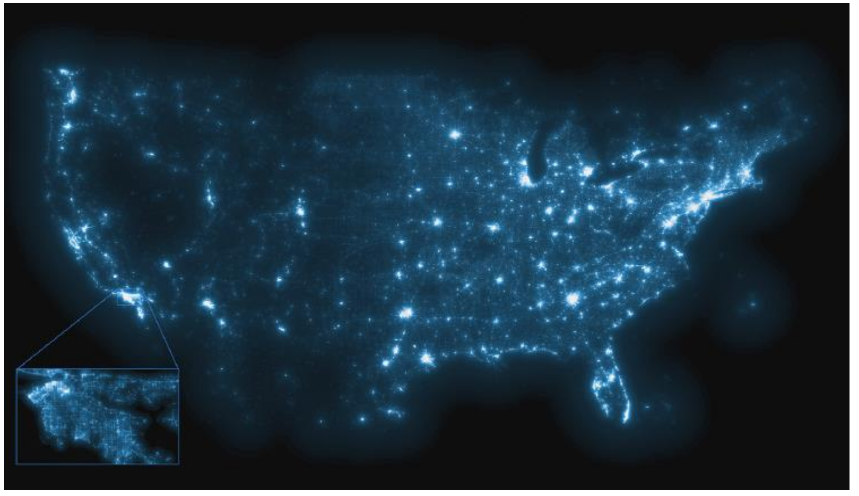
Regional sentiment about Palestinian suicide bombing (blue, normalized) and bombing frequency (red, normalized) are correlated, with sentiment leading bombing events by 12 months ($R = 0.8$, $p < 0.0001$).



Case Study: H1N1 Epidemic

Introduction

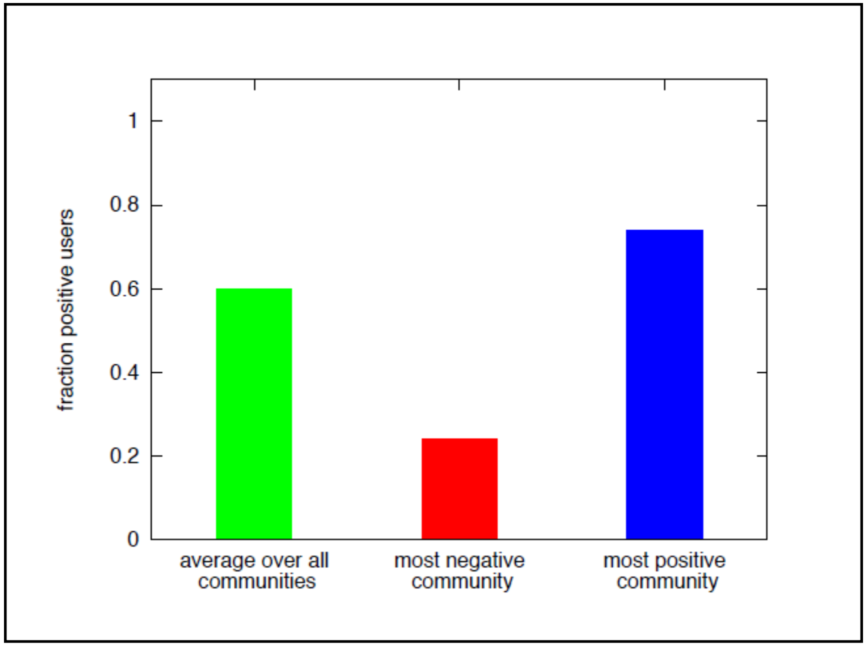
- Public sentiment regarding vaccination can strongly affect vaccination rates [Signorini et al. 2011, Bhattacharya et al. 2012] and therefore the risk of disease outbreak and overall security of a population.
- We estimate public sentiment concerning vaccination for H1N1 flu during the second half of 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. This fact could have important implications for vaccination campaigns.



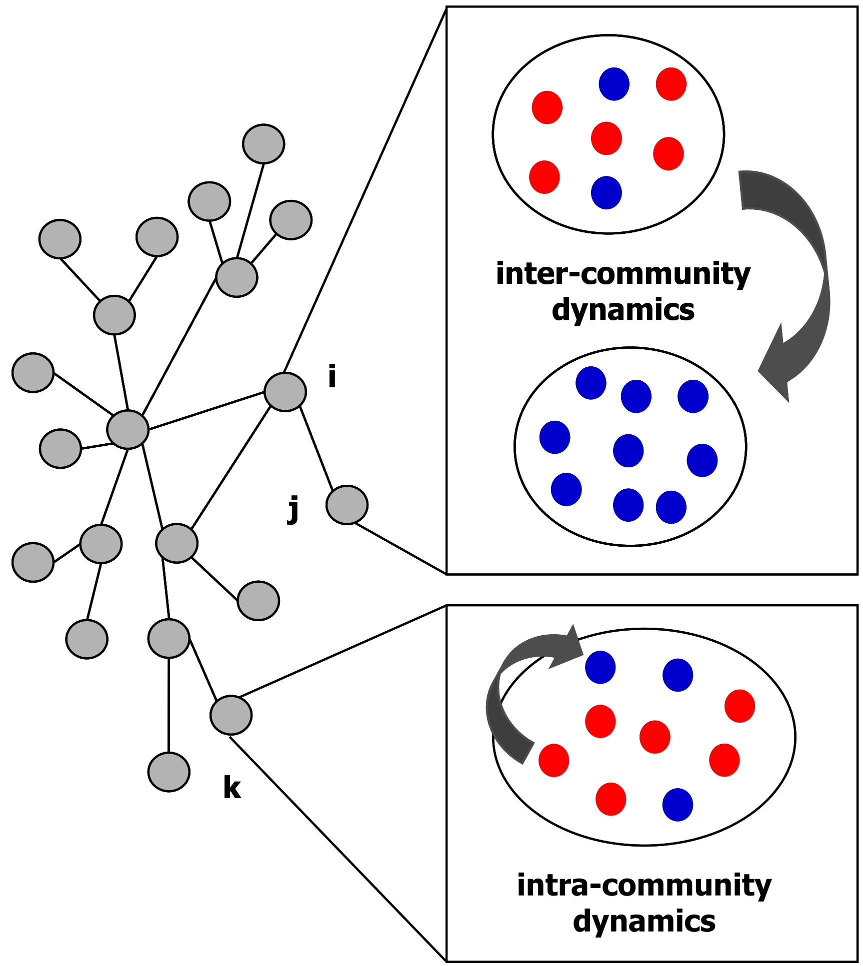
Case Study: H1N1 Epidemic

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:

- agent dynamics – competing opinions model [Bettencourt et al. 2006, Colbaugh/Glass 2012];
- social network – family of multiple-community graphs [Candia/Mazzitello 2008].



Case Study: H1N1 Epidemic

Implication of heterogeneous sentiment distribution (cont'd)

Sample results

<u>Sample Simulation Results</u>

1. Baseline:

balanced seeds, homog. communities yields balanced outcomes (e.g. 50/50 → 950/950).

2. Unbalanced seeds, homog. communities:

outcomes are proportional to the (initial) seed imbalance (e.g. 50/10 → 1500/300).

3. Unbalanced seeds, heterog. communities:

outcomes can be much more balanced than the initialization (e.g. 50/10 → 1000/750).