Overcoming Language Variation in Sentiment Analysis with Social Attention

Yi Yang
Bloomberg LP

Work performed at Georgia Tech with Jacob Eisenstein.
Language variation in sentiment analysis

“I would like to believe he’s sick rather than just mean and evil.”
Language variation in sentiment analysis

“I would like to believe he’s sick rather than just mean and evil.”

“You could’ve been getting down to this sick beat.”
Language variation in sentiment analysis

I am sick and weak

THIS IS SO SICK THANK U
Language variation in sentiment analysis

I am sick and weak

THIS IS SO SICK THANK U
Language variation in sentiment analysis

I am sick and weak

THIS IS SO SICK THANK U

ALEX THIS IS SO SICK
Language variation in sentiment analysis

I am sick and weak

THIS IS SO SICK THANK U

ALEX THIS IS SO SICK

<table>
<thead>
<tr>
<th></th>
<th>Reviews</th>
<th>Tweets</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>80.2</td>
<td>71.2</td>
</tr>
</tbody>
</table>

F1
Personalized sentiment analysis

- **Goal**: personalized conditional likelihood, \( p(y|x, a) \).
- \( x \) is the text, and \( a \) is the author.
Personalized sentiment analysis

- **Goal**: personalized conditional likelihood, \( p(y|x, a) \).
- \( x \) is the text, and \( a \) is the author.
Personalized sentiment analysis

- **Goal**: personalized conditional likelihood, \( p(y \mid x, a) \).
- **Problem**: we have labeled examples for only a few authors.

\[ p(y \mid x, a) \]

\[ x \text{ is the text, and } a \text{ is the author.} \]
Homophily to the rescue?

**Homophily**: neighbors have similar properties.

Thelwall (2009); Al Zamal et al. (2012)
Homophily to the rescue?

**Homophily:** neighbors have similar properties.

Thelwall (2009); Al Zamal et al. (2012)
Evidence for linguistic homophily

**Pilot study**: is classifier accuracy *assortative* on the Twitter social network?
Evidence for linguistic homophily

**Pilot study**: is classifier accuracy * assortative* on the Twitter social network?

\[
\text{assort}(G) = \frac{1}{\#|G|} \sum_{(i,j) \in G} \delta(y_i = \hat{y}_i)\delta(y_j = \hat{y}_j) + \delta(y_i \neq \hat{y}_i)\delta(y_j \neq \hat{y}_j)
\]

- Whether a sentiment classifier tends to make **consistent** predictions for social neighbors.
Evidence for linguistic homophily

Network rewiring: degree-preserving randomization
Evidence for linguistic homophily

Network rewiring: degree-preserving randomization

A -- B

C -- D
Evidence for linguistic homophily

Network rewiring: degree-preserving randomization
Evidence for linguistic homophily

**Network rewiring**: degree-preserving randomization

![Graphs showing network rewiring](image)

- **Follow**: Assortativity over rewiring epochs for original and random rewiring networks.
- **Mention**: Similar pattern as follow but with different values.
- **Retweet**: Assortativity for retweet interactions showing the effect of rewiring.

In each graph, the red dashed line represents the original network's assortativity, while the blue line shows the assortativity of the random rewiring network.
Model
Personalization by ensemble

\[ p(y|x, a) = \sum_{k=1}^{K} \Pr(Z_a = k|a, G) \times p(y|x, Z_a = k) \]

ensemble weights

basis models
Personalization by ensemble

\[ p(y|x, a) = \sum_{k=1}^{K} Pr(Z_a = k|a, G) \times p(y|x, Z_a = k) \]

- Train each basis model with all the labeled data.
  - Employ ConvNets as basis models.
Personalization by ensemble

\[ p(y|x, a) = \sum_{k=1}^{K} \underbrace{Pr(Z_a = k|a, G)} \times \underbrace{p(y|x, Z_a = k)} \]

- Train each basis model with all the labeled data.
  - Employ ConvNets as basis models.
- Apply linguistic homophily:
  - Adopt similar ensemble weights for social neighbors.
  - De-correlate errors made by different basis models.
Network-driven personalization

- For each author, estimate a node embedding $v_a$ (Tang et al., 2015).
- Nodes who share neighbors get similar embeddings.
Network-driven personalization

- For each author, estimate a node embedding $v_a$ (Tang et al., 2015).
- Nodes who share neighbors get similar embeddings.

Social attention:

$$Pr(Z_a = k|a, G) = \text{SoftMax}(f(v_a))$$
Learning

- Jointly train with cross-entropy loss:

\[
\ell(\Theta) = -\sum_{t=1}^{T} 1[Y^* = t] \log \Pr(Y = t \mid x, a)
\]
Jointly train with cross-entropy loss:

$$\ell(\Theta) = - \sum_{t=1}^{T} 1[Y^* = t] \log \Pr(Y = t \mid x, a)$$

Problem: network information tends to be ignored.
Learning

- Jointly train with cross-entropy loss:
  \[ \ell(\Theta) = - \sum_{t=1}^{T} 1[Y^* = t] \log \Pr(Y = t \mid x, a) \]

  **Problem:** network information tends to be ignored.

- Pre-train basis models with instance-weighted losses:
  \[ \ell_k = -\alpha_{a,k} \sum_{t=1}^{T} 1[Y^* = t] \log \Pr(Y = t \mid x, Z_a = k) \]
Experiments
Data

- SemEval Twitter sentiment analysis data.
  - 18,024 tweets
  - Follow, mention, retweet networks
  - Network expansion

Nakov et al. (2013); Rosenthal et al. (2015); Tang et al. (2012)
Data

- SemEval Twitter sentiment analysis data.
  - 18,024 tweets
  - Follow, mention, retweet networks
  - Network expansion

- Ciao product review sentiment analysis data.
  - 100,000 reviews
  - User trust network

Nakov et al. (2013); Rosenthal et al. (2015); Tang et al. (2012)
Results: SemEval Twitter data

Our implementation

- Mixture of experts
- Random attention
- Social attention

Best published results

NCNN baseline
F1: 68.4

+ 1.9

NLSE

Astudillo et al. (2015)
Results: SemEval Twitter data

Our implementation

- Mixture of experts: + 0.0
- Random attention: + 0.1
- Social attention: + 1.9

Best published results: + 1.9

CNN baseline: F1: 68.4

Astudillo et al. (2015)
Results: SemEval Twitter data

Our implementation

- Mixture of experts: + 0.0
- Random attention: + 0.1
- Social attention: + 2.8

Best published results

- NLSE: + 1.9

CNN baseline

F1: 68.4

Astudillo et al. (2015)
### Variable sentiment words

<table>
<thead>
<tr>
<th>More positive</th>
<th>More negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  bang</td>
<td>dear</td>
</tr>
<tr>
<td>2  chilling</td>
<td>like</td>
</tr>
<tr>
<td>3  ass</td>
<td>trust</td>
</tr>
<tr>
<td>4  insane</td>
<td>wealth</td>
</tr>
<tr>
<td>5  ruin</td>
<td>strong</td>
</tr>
<tr>
<td>bang</td>
<td>like</td>
</tr>
<tr>
<td>loss</td>
<td>god</td>
</tr>
<tr>
<td>fever</td>
<td>yeah</td>
</tr>
<tr>
<td>broken</td>
<td>wow</td>
</tr>
<tr>
<td>fucking</td>
<td></td>
</tr>
<tr>
<td>cold</td>
<td>satisfy</td>
</tr>
<tr>
<td>ill</td>
<td>trust</td>
</tr>
<tr>
<td>sick</td>
<td>wealth</td>
</tr>
<tr>
<td>suck</td>
<td>strong</td>
</tr>
<tr>
<td></td>
<td>lmao</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>ass</td>
<td>talent</td>
</tr>
<tr>
<td>damn</td>
<td>honestly</td>
</tr>
<tr>
<td>piss</td>
<td>voting</td>
</tr>
<tr>
<td>bitch</td>
<td>win</td>
</tr>
<tr>
<td>shit</td>
<td>clever</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>insane</td>
<td>lmao</td>
</tr>
<tr>
<td>bawling</td>
<td>super</td>
</tr>
<tr>
<td>fever</td>
<td>lol</td>
</tr>
<tr>
<td>weird</td>
<td>haha</td>
</tr>
<tr>
<td>cry</td>
<td>hahaha</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>ruin</td>
<td>lovatics</td>
</tr>
<tr>
<td>silly</td>
<td>wish</td>
</tr>
<tr>
<td>bad</td>
<td>beliebers</td>
</tr>
<tr>
<td>boring</td>
<td>arianators</td>
</tr>
<tr>
<td>dreadful</td>
<td>kendall</td>
</tr>
</tbody>
</table>
## Results: Ciao review data

<table>
<thead>
<tr>
<th></th>
<th>+ 0.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN baseline</td>
<td>F1: 74.4</td>
</tr>
<tr>
<td>Mixture of</td>
<td></td>
</tr>
<tr>
<td>experts</td>
<td></td>
</tr>
<tr>
<td>Random</td>
<td></td>
</tr>
<tr>
<td>attention</td>
<td></td>
</tr>
<tr>
<td>Social</td>
<td></td>
</tr>
<tr>
<td>attention</td>
<td></td>
</tr>
</tbody>
</table>

Tang et al. (2012)
Results: Ciao review data

CNN baseline
F1: 74.4

+ 0.0
Mixture of experts
Random attention
Social attention

+ 1.0

Tang et al. (2012)
Results: Ciao review data

Mixture of experts + 1.0
Random attention + 0.0
Social attention + 1.8

CNN baseline F1: 74.4
Tang et al. (2012)
Conclusions and future work

- Language variation poses challenges in sentiment analysis.
- Linguistic homopily alleviates the data sparsity issue for estimating personalized models.
- Social attention mechanism significantly improves accuracy.
- The socially-infused ensemble architecture can be applied to other tasks such as tagging, parsing, etc.