Seeing stars when there aren't many stars

Graph-based semi-supervised learning for sentiment categorization

Andrew B. Goldberg, Xiaojin “Jerry” Zhu
Computer Sciences Department
University of Wisconsin-Madison
Goodbye Academia, Hello Hollywood!

Not quite!
Sentiment Categorization

“...captivating... special effects were amazing...”

“...excellent acting... quite good and believable...”

“...weak, lame attempts at humor...bland as can be...”
“...captivating... special effects were amazing...”

“...excellent acting... quite good and believable...”

“...weak, lame attempts at humor...bland as can be...”
Graph-Based Learning

- This is rating inference [Pang+Lee05]
- We use graph-based semi-supervised learning
- Main assumption encoded in graph: Similar documents should have similar ratings

Graph-Based Learning

**Labeled**

- “…captivating… special effects were amazing…” ★★★★★
- “…excellent acting… quite good and believable…” ★★★★☆
- “…weak, lame attempts at humor… bland as can be…” ★☆☆☆☆

**Unlabeled**

- “…captivating piece of work from an excellent team…”
- “…excellent acting makes up for bland, lame scenery…”
- “…preview quite good, but acting was weak, way off…”
- “…weak, bad… pathetically lame… acting way off…”
Graph-Based Learning

Labeled

“...captivating... special effects were amazing...”
★★★★★

“...excellent acting... quite good and believable...”
★★★★☆

“...weak, lame attempts at humor... bland as can be...”
★☆☆☆☆

Unlabeled

“...captivating piece of work from an excellent team...”
★★★★☆

“...excellent acting makes up for bland, lame scenery...”
★★★★☆

“...preview quite good, but acting was weak, way off...”
★★☆☆☆

“...weak, bad... pathetically lame... acting way off...”
★☆☆☆☆

Truth

★★★★★
★★★★☆
★★☆☆☆
★☆☆☆☆
Graph-Based Learning

Labeled

“...captivating... special effects were amazing...”
★★★★★

“...excellent acting... quite good and believable...”
★★★☆☆

“...weak, lame attempts at humor... bland as can be...”
★★☆☆☆

Unlabeled

“...captivating piece of work from an excellent team...”

“...excellent acting makes up for bland, lame scenery...”

“...preview quite good, but acting was weak, way off...”

“...weak, bad... pathetically lame... acting way off...”

Truth
★★★★
★★★★
★★★☆
★★☆☆
★★☆☆
Supervised
★★★★
★★★☆
★★★☆
★★★☆
★★★☆
Graph-Based Learning

Labeled

“...captivating... special effects were amazing...”
★★★★★

“...excellent acting... quite good and believable...”
★★★★☆

“...weak, lame attempts at humor... bland as can be...”
★★☆☆☆

Unlabeled

“...captivating piece of work from an excellent team...”
★★★★★

“...excellent acting makes up for bland, lame scenery...”
★★★★☆

“...preview quite good, but acting was weak, way off...”
★★★★☆

“...weak, bad... pathetically lame... acting way off...”
★★★★☆

Truth

★★★★★  ★★★★★ ✔
★★★★☆  ★★★★★ ✗
★★★★☆  ★★★★★
★★★★☆  ★★★★★

Supervised
Graph-Based Learning

**Labeled**

- “...captivating... special effects were amazing...”
  - ★★★★★

- “...excellent acting... quite good and believable...”
  - ★★★★☆

- “...weak, lame attempts at humor... bland as can be...”
  - ★☆☆☆☆

**Unlabeled**

- “...captivating piece of work from an excellent team...”
  - ★★★★★

- “...excellent acting makes up for bland, lame scenery...”
  - ★★★☆☆

- “...preview quite good, but acting was weak, way off...”
  - ★★★☆☆

- “...weak, bad... pathetically lame... acting way off...”
  - ★☆☆☆☆

**Truth**

- ★★★★★
- ★★★☆☆
- ★★★☆★
- ★☆☆☆☆

**Supervised**

- ✔ ✔ ✔ ✔
- ✗ ✔ ✔ ✗
- ✗ ✗ ✔ ✗
- ✔ ✗ ✔ ✔
50% Accuracy 😞

Labeled

"...captivating... special effects were amazing..."
★★★★

"...excellent acting... quite good and believable..."
★★★★☆

"...weak, lame attempts at humor... bland as can be..."
★☆☆☆☆

Unlabeled

"...captivating piece of work from an excellent team..."
★★★★

"...excellent acting makes up for bland, lame scenery..."
★★★★☆

"...preview quite good, but acting was weak, way off..."
★★★★☆

"...weak, bad... pathetically lame... acting way off..."
★★★★☆

Truth

★ ★ ★ ★ ★
★ ★ ★ ★ ☆
★ ★ ★ ★ ☆
★ ★ ★ ★ ☆
★ ★ ★ ★ ☆

Supervised

✔
✘
✘
✔
✔
Graph-Based Learning

Labeled

“...captivating... special effects were amazing...”
★★★★★

“...excellent acting... quite good and believable...”
★★★★☆

“...weak, lame attempts at humor... bland as can be...”
★★☆☆☆

Unlabeled

“...captivating piece of work from an excellent team...”
★★★★★

“...excellent acting makes up for bland, lame scenery...”
★★★★☆

“...preview quite good, but acting was weak, way off...”
★★★★☆

“...weak, bad... pathetically lame... acting way off...”
★★★★☆

Truth

Semi-Supervised

★★★★★

★★★★☆

★★★★☆

★★★★☆
Graph-Based Learning

Labeled

“...captivating... special effects were amazing...”
★ ★ ★ ★ ★

“...excellent acting... quite good and believable...”
★ ★ ★ ★ ★

“...weak, lame attempts at humor... bland as can be...”
★ ★ ★ ★ ★

Unlabeled

“...captivating piece of work from an excellent team...”

“...excellent acting makes up for bland, lame scenery...”

“...preview quite good, but acting was weak, way off...”

“...weak, bad... pathetically lame... acting way off...”

Truth

Semi-Supervised

★ ★ ★ ★ ★

★ ★ ★ ★ ★

★ ★ ★ ★ ★

★ ★ ★ ★ ★

★ ★ ★ ★ ★

★ ★ ★ ★ ★

★ ★ ★ ★ ★

★ ★ ★ ★ ★
Graph-Based Learning

100% Accuracy 😊

Labeled

“...captivating... special effects were amazing...”
★★★★★

“...excellent acting... quite good and believable...”
★★★★☆

“...weak, lame attempts at humor... bland as can be...”
★★☆☆☆

Unlabeled

“...captivating piece of work from an excellent team...”

“...excellent acting makes up for bland, lame scenery...”

“...preview quite good, but acting was weak, way off...”

“...weak, bad... pathetically lame... acting way off...”

Truth

Semi-Supervised

★★★★★  ★★★★★

★★★★☆  ★★★★★

★★☆☆☆  ★★★★★

★★☆☆☆  ★★★★★
How it really works
Goal

- Assign a discrete numeric rating \( f(x) \) to each document \( x \)

\[
f : \begin{array}{c}
\text{document} \\
\end{array} \rightarrow \begin{cases}
\text{☆☆☆☆☆} \\
\text{☆☆☆☆☆} \\
\text{☆☆☆☆☆} \\
\text{☆☆☆☆☆} 
\end{cases}
\]
Our Approach

• Get initial predictions using SVM
• Improve predictions using graph-based SSL
  – Nodes = reviews
  – Edges = assumed relations between reviews
  – Find the optimal $f$ over the graph
Measuring Loss over the Graph

Similar reviews should get similar ratings

Loss over this edge = \(w \cdot (f_A - f_B)^2\)

Task: Minimize loss \(L(f)\) over whole graph
## Semi-Supervised Learning Graph

<table>
<thead>
<tr>
<th>Unlabeled</th>
<th>Labeled</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Unlabeled Nodes" /></td>
<td><img src="image2.png" alt="Labeled Nodes" /></td>
</tr>
</tbody>
</table>
Semi-Supervised Learning Graph

Unlabeled

Labeled

Loss $L(f) =$
Semi-Supervised Learning Graph

Unlabeled

\[ \sum_{i \in U} (f_i - \hat{y}_i)^2 \]

Labeled

Loss \( L(f) = \)
Semi-Supervised Learning Graph

Unlabeled

\[ \sum_{i \in U} (f_i - \hat{y}_i)^2 + \sum_{i \in L} M(f_i - y_i)^2 \]

Labeled

Given label \( y_i \)

Predicted label \( \hat{y}_i \)
Semi-Supervised Learning Graph

Unlabeled

Predicted label $\hat{y}_i$

Given label $y_i$

Labeled

Loss $L(f) =$

$$\sum_{i \in U} (f_i - \hat{y}_i)^2 + \sum_{i \in L} M(f_i - y_i)^2$$

Minimization is fairly trivial $f_i = \hat{y}_i$ or $y_i$
Semi-Supervised Learning Graph

\[ L(f) = \sum_{i \in U} (f_i - \hat{y}_i)^2 + \sum_{i \in L} M(f_i - y_i)^2 + \sum_{i \in U, j \in k\text{NN}_L(i)} a \cdot w_{ij} (f_i - f_j)^2 \]
Semi-Supervised Learning Graph

Loss $L(f) =$

\[ \sum_{i \in U} (f_i - \hat{y}_i)^2 + \sum_{i \in L} M(f_i - y_i)^2 + \sum_{i \in U} \sum_{j \in kNN_L(i)} aw_{ij}(f_i - f_j)^2 + \sum_{i \in U} \sum_{j \in k'NN_U(i)} bw_{ij}(f_i - f_j)^2 \]
Minimization is now \textit{non-trivial}

Loss $L(f) = \sum_{i \in U} (f_i - \hat{y}_i)^2 + \sum_{i \in L} M(f_i - y_i)^2 + \sum_{i \in U} \sum_{j \in k \text{NN}_L(i)} a w_{ij} (f_i - f_j)^2 + \sum_{i \in U} \sum_{j \in k' \text{NN}_U(i)} b w_{ij} (f_i - f_j)^2$
Finding a Closed-Form Solution

\[ L(f) = \sum_{i \in U} (f_i - \hat{y}_i)^2 + \sum_{i \in L} M(f_i - y_i)^2 + \]
\[ \sum_{i \in U} a w_{ij} (f_i - f_j)^2 + \sum_{i \in U} b w_{ij} (f_i - f_j)^2 \]
Finding a Closed-Form Solution

\[ L(f) = \sum_{i \in U} (f_i - \hat{y}_i)^2 + \sum_{i \in L} M (f_i - y_i)^2 + \]
\[ \sum_{i \in U} \sum_{j \in kNN_L(i)} aw_{ij} (f_i - f_j)^2 + \sum_{i \in U} \sum_{j \in k' NN_U(i)} bw_{ij} (f_i - f_j)^2 \]

Can rewrite \( L(f) \) in matrix notation as:

\[ (f - y)^\top C (f - y) + \eta f^\top \Delta f \]
Finding a Closed-Form Solution

Can rewrite $L(f)$ in matrix notation as:

$$(f - y)^T C (f - y) + \eta f^T \Delta f$$
Finding a Closed-Form Solution

Can rewrite $L(f)$ in matrix notation as:

$$\left(f - y\right)^T C \left(f - y\right) + \eta f^T \Delta f$$

Vector of $f$ values for all reviews

Vector of given labels $y_i$ for labeled reviews and predicted labels $\hat{y}_i$ for unlabeled reviews

Matrix $C$:

- Labeled
- Unlabeled

All other matrix entries are 0

1

\[ C = \begin{pmatrix}
1 & & & & \\
& 1 & & & \\
& & 1 & & \\
& & & \ddots & \\
& & & & 1
\end{pmatrix} \]
Finding a Closed-Form Solution

Can rewrite $L(f)$ in matrix notation as:

$$(f - y)^T C (f - y) + \eta f^T \Delta f$$

[Diagram showing a graph with nodes labeled $\hat{y}_i$, $y_i$, $M$, and edges labeled $b \cdot w_{ij}$ and $a \cdot w_{ij}$]
Graph Laplacian Matrix

Assume $n$ labeled and unlabeled documents

$W = n \times n$ weight matrix

$D = n \times n$ diagonal degree matrix, where

$$D_{ii} = \sum_{j=1}^{n} W_{ij}$$

Graph Laplacian matrix is

$$\Delta = D - W$$
Finding a Closed-Form Solution

• $L(f)$ in matrix notation:

$$(f - y)^T C (f - y) + \eta f^T \Delta f$$

• Set gradient to zero and solve for $f$:

$$f = (C + \eta \Delta)^{-1} C y$$
Experiments

- Task: Predict 1 to 4 star ratings for reviews
  - 4-author data used by [Pang+Lee05]
  - Predicted \( \hat{y}_i \) values with SVM\textsuperscript{light} regression using \{0,1\} word vectors
  - Positive-sentence percentage (PSP) similarity [Pang+Lee05]
  - Tuned parameters with cross validation
Results

Author A

Average test set accuracy

log labeled set size
Results

```plaintext
Author A

Average test set accuracy

SVM regression
Metric labeling

log labeled set size

1 2 3 4 5 6 7 8

0.65
0.6
0.55
0.5
0.45
0.4
0.35
0.4
0.35
0.4
0.35
```
Results
Results
Results

Graph-based SSL outperforms other methods for small labeled set sizes.
Conclusions

• Adapted graph-based semi-supervised learning to sentiment analysis domain

• Designed a graph for rating inference

• Showed benefit of SSL using movie review data

Thank you! Any questions?