Social Influence and Sentiment Analysis
—From Sentiment to Emotion Analysis in Social Networks

Jie Tang

Department of Computer Science and Technology
Tsinghua University, China
Networked World

- **Facebook**
  - 1.65 billion MAU (users)
  - 2.5 trillion minutes/month

- **Twitter**
  - 255 million MAU
  - Peak: 143K tweets/s

- **WeChat**
  - QQ: 800 million MAU
  - WeChat: 700 million MAU

- **Amazon**
  - 304 million active users
  - 14 billion items/year

- **Alibaba Group**
  - 710 million trans. on 11/11
  - 13.6 billion USD in 24 hrs

- **Influencing our daily life**
The Era of Big Social Data

• We generate $2.5 \times 10^{18}$ byte big data per day.

• Big social data:
  – 90% of the data was generated in the past 2 yrs
  – Mining in single data center → mining deep knowledge from multiple data sources
User Opinion and Influence: “Love Obama”

I love Obama

Obama is fantastic

I hate Obama, the worst president ever

Obama is great!

No Obama in 2012!

He cannot be the next president!

Positive  Negative
Does Social Influence really matter?

- **Case 1:** Social influence and political mobilization[1]
  - Will online political mobilization really work?

A controlled trial (with 61M users on FB)

- **Social msg group:** was shown with msg that indicates one’s friends who have made the votes.
- **Informational msg group:** was shown with msg that indicates how many other.
- **Control group:** did not receive any msg.

Case 1: Social Influence and Political Mobilization

Social msg group v.s. Info msg group

Result: The former were 2.08% (t-test, $P<0.01$) more likely to click on the “I Voted” button.

Social msg group v.s. Control group

Result: The former were 0.39% (t-test, $P=0.02$) more likely to actually vote (via examination of public voting records).

Twitter Data

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

Table 1: Statistics for our main datasets.

<table>
<thead>
<tr>
<th>Topic</th>
<th># users</th>
<th>#t-follow edges</th>
<th># on-topic tweets</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>dir.</td>
<td>mutual</td>
</tr>
<tr>
<td>Obama</td>
<td>889</td>
<td>7,838</td>
<td>2,949</td>
</tr>
<tr>
<td>Sarah Palin</td>
<td>310</td>
<td>1,003</td>
<td>264</td>
</tr>
<tr>
<td>Glenn Beck</td>
<td>313</td>
<td>486</td>
<td>159</td>
</tr>
<tr>
<td>Lakers</td>
<td>640</td>
<td>2,297</td>
<td>353</td>
</tr>
<tr>
<td>Fox News</td>
<td>231</td>
<td>130</td>
<td>32</td>
</tr>
</tbody>
</table>

From text sentiment to user sentiment

Obama is making the repubs look silly and petty.

However, the social text is really short and noisy …

Only thing we have to fear is Obama himself & Pelosi & Cong & liberal news & Dems &...

Barack Obama can no more disown ACORN than he could disown his own grandmother.
From user sentiment to network sentiment

① Who influenced who? What is the influence probability?

② Can we leverage the social influence to help sentiment analysis?

I love Obama
Obama is fantastic
Obama is great!
I hate Obama, the worst president ever
No Obama in 2012!
He cannot be the next president!

Positive
Negative

0.3
0.2
0.5
0.4
0.05
0.7
0.74
0.1
0.1
0.5
0.05

Can we leverage the social influence to help sentiment analysis?
Sentiment Influence in Twitter

Shared sentiment conditioned on type of connection.
—people tend to follow the opinion of their friends

Selection

Connectedness conditioned on labels
—people tend to create relationships with other people who share the same opinion with them
Learning for network sentiment analysis

Networked Classification Model: Learning for sentiment analysis by considering the network information

Another challenge: labeled data is very limited…
Semi-supervised Factor Graph Model

**Semi-FGM**: learning to classify sentiments by considering both content and network structure in a semi-supervised fashion.

$$f_{k,\ell}(y_v, \hat{y}) = \begin{cases} \frac{w_{\text{labeled}}}{\text{tweets}_{v_i}} & y_v = k, \hat{y} = \ell, v_i \text{ labeled} \\ \frac{w_{\text{unlabeled}}}{\text{tweets}_{v_i}} & y_v = k, \hat{y} = \ell, v_i \text{ unlabeled} \\ 0 & \text{otherwise} \end{cases}$$

Semi-FGM: learning to classify sentiments by considering both content and network structure in a semi-supervised fashion.

\[ h_{k,\ell}(y_v, y_{v'}) = \begin{cases} \frac{w_{\text{relation}}}{|\text{Neighbors}_{v_i}|} & y_v = k, y_{v'} = \ell \\ 0 & \text{otherwise} \end{cases} \]

Semi-supervised Factor Graph Model

\[
\log P(Y) = \left( \sum_{v_i \in V} \left[ \sum_{t \in \text{tweets}_{v_i}, k, \ell} \mu_{k, \ell} f_{k, \ell}(y_{v_i}, \hat{y}_t) \right] + \sum_{v_j \in \text{Neighbors}_{v_i}, k, \ell} \lambda_{k, \ell} h_{k, \ell}(y_{v_i}, y_{v_j}) \right) - \log Z,
\]

\[
f_{k, \ell}(y_{v_i}, \hat{y}_t) = \begin{cases} \frac{w_{\text{labeled}}}{|\text{tweets}_{v_i}|} & y_{v_i} = k, \hat{y}_t = \ell, v_i \text{ labeled} \\ \frac{w_{\text{unlabeled}}}{|\text{tweets}_{v_i}|} & y_{v_i} = k, \hat{y}_t = \ell, v_i \text{ unlabeled} \\ 0 & \text{otherwise} \end{cases}
\]

\[
h_{k, \ell}(y_{v_i}, y_{v_j}) = \begin{cases} \frac{w_{\text{relation}}}{|\text{Neighbors}_{v_i}|} & y_{v_i} = k, y_{v_j} = \ell \\ 0 & \text{otherwise} \end{cases}
\]
Parameter Estimation for Semi-FGM

• "NoLearning": simply use counts from the labeled subset of the data

\[
\lambda_{k,\ell} := \frac{\sum_{(v_i,v_j) \in E_{\text{labeled}}} I(y_{v_i} = k, y_{v_j} = \ell)}{\sum_{(v_i,v_j) \in E_{\text{labeled}}} I(y_{v_i} = k, y_{v_j} = 1) + I(y_{v_i} = k, y_{v_j} = 0)}
\]

the subset of edges in our dataset in which both endpoints are labeled

indicator function

• SampleRank ("Learning"): A sampling-based learning algorithm using Metropolis–Hastings
SampleRank ("Learning")

Input: Heterogeneous graph $HG$ with labels on some of the user nodes, learning rate $\eta$

Output: Parameter values $\phi$ and full label-vector $Y$

Randomly initialize $Y$;
Initialize $\phi$ from NoLearning;

for $i := 1$ to Number of Steps do
  $Y_{\text{new}} := \text{Sample}(Y)$;
  if ($\text{RelPerf}(Y_{\text{new}}, Y) > 0$ and $\text{LLR}_\phi(Y_{\text{new}}, Y) < 0$) // performance is better but the objective function is lower
    or ($\text{RelPerf}(Y_{\text{new}}, Y) < 0$ and $\text{LLR}_\phi(Y_{\text{new}}, Y) > 0$) // performance is worse but the objective function is higher
    then
      $\phi := \phi - \eta \nabla \phi \text{LLR}_\phi(Y_{\text{new}}, Y)$;
    end
  if convergence then
    break;
  end
  if $\text{RelPerf}(Y_{\text{new}}, Y) > 0$ then
    $Y := Y_{\text{new}}$;
  end
end

likelihood ratio of new sample $Y_{\text{new}}$ and previous label $Y$ for all users

Update model parameters when two results are inconsistent

Relative performance between new sample $Y_{\text{new}}$ and previous label $Y$ on labeled user only.
Results of network sentiment analysis

- **Twitter**
  - 1,414,340 users and 480,435,500 tweets
  - 274,644,047 t-follow edges and 58,387,964 @ edges

- **Methods**
  - SVM Vote
  - Semi-FGM (NoLearning)
  - Semi-FGM (SampleRank)

- **Measures**
  - Accuracy and Macro F1
<table>
<thead>
<tr>
<th>User ID</th>
<th>SVM Vote</th>
<th>HGM</th>
<th>True</th>
<th>Tweet</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NEG</td>
<td>POS</td>
<td>POS</td>
<td>Obama is making the repubs look silly and petty. #hrc</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Is happy Obama is President</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Obama collectable <a href="http://tinyurl.com/c5u7jf">http://tinyurl.com/c5u7jf</a></td>
</tr>
<tr>
<td>2</td>
<td>NEG</td>
<td>POS</td>
<td>POS</td>
<td>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!!</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>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!</td>
</tr>
<tr>
<td>3</td>
<td>NEG</td>
<td>POS</td>
<td>POS</td>
<td>RT @TeaPartyProtest Only thing we have 2 fear is Obama himself &amp; Pelosi &amp; Cong &amp; liberal news &amp; Dems &amp;… <a href="http://ow.ly/15M9Xv">http://ow.ly/15M9Xv</a></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RT @GlennBeckClips: Barack Obama can no more disown ACORN than he could disown his own grandmother. #TCOT</td>
</tr>
<tr>
<td>4</td>
<td>POS</td>
<td>NEG</td>
<td>NEG</td>
<td>RT @JosephAGallant Twitlonger: Suppose I wanted to Immigrant to Mexico? A Letter to President Obama. <a href="http://tl.gd/1kr5rh">http://tl.gd/1kr5rh</a></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>George Bush was and acted like a war time President. Obama is on a four year power grab and photo op. #tcot</td>
</tr>
<tr>
<td>5</td>
<td>POS</td>
<td>NEG</td>
<td>NEG</td>
<td>ObamaCare forces Americans to buy or face a fine! It is UNCONSTITUTIONAL to force us to buy obamacare. Marxist Govt. taking our Freedoms!</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Look up Chicago Climate Exchange, an organization formed years ago by Obama &amp; his Marxist-Commie Cronies to form a profit off cap &amp; trade.</td>
</tr>
<tr>
<td>6</td>
<td>POS</td>
<td>NEG</td>
<td>NEG</td>
<td></td>
</tr>
</tbody>
</table>
Performance

Accuracy

MacroF1
Performance Analysis in Different Topics
Results of Different Learning Algorithms

- Obama
- Sarah Palin
- Glenn Beck
- Lakers
- Fox News
# Twitter to Weibo

![China Map](image)

## Guangzhou Happiness Index

<table>
<thead>
<tr>
<th>Province</th>
<th>Happiness Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shanghai</td>
<td>40.76491</td>
</tr>
<tr>
<td>Beijing</td>
<td>40.27117</td>
</tr>
<tr>
<td>Tianjin</td>
<td>33.05676</td>
</tr>
<tr>
<td>Jiangsu</td>
<td>32.77226</td>
</tr>
<tr>
<td>Hunan</td>
<td>32.52948</td>
</tr>
<tr>
<td>Zhejiang</td>
<td>32.24124</td>
</tr>
<tr>
<td>Sichuan</td>
<td>31.95578</td>
</tr>
</tbody>
</table>

Using this photo best //@我爱黄世仁-奎：好。

August 04, 2021 15:36

Invoked: 332

Replies: 0 | Comments: 3

I know it is my own fault, so this is the result, but why don't I give you the reward I promised? I always strive to be a good person, why can't I be good?

March 13, 2021 21:49

Invoked: 4

Replies: 0 | Comments: 0

I participated in the favorable dormitory - pleasant feelings发起的投票【最佳展示宿舍】，我投给了“生技公寓2519 神迹519”这个选项，你也快来表态吧：http://t.cn/zOuIWlq

May 21, 2021 23:42

Invoked: 28

Replies: 0 | Comments: 1
We have a picture of sentiment analysis in social networks…

- From text sentiment to user sentiment
- From user sentiment to network sentiment
- Challenges:
  - Short text and noisy data
  - Limited labeled data
  - Networked user sentiments
- Proposal of a Semi-supervised Factor Graph Model (Semi-FGM) to learn to classify sentiments by considering both content and network structure
Now, let us think…

• What are the fundamental factors behind
  – What is behind the network of social users?
  – What is behind the sentiment of social users?
Well, what is the fundamental factor…

Info. Space vs. Social Space

From the social network research perspective, what are the fundamental factors behind?

Understanding the mechanism of interaction dynamics
Topic-based Social Influence Analysis

<table>
<thead>
<tr>
<th>Topics</th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Politics</td>
<td>0.3</td>
<td>0.7</td>
</tr>
<tr>
<td>Entertainment</td>
<td>0.2</td>
<td>0.1</td>
</tr>
<tr>
<td>Trademarks</td>
<td>0.5</td>
<td>0.74</td>
</tr>
</tbody>
</table>

How to?

Politics

Politics
The Solution: Topical Affinity Propagation

Basic Idea:
If a user is located in the center of a “Market” community, and is “similar” to the other users, then she/he would have a strong influence on the other users.

—Homophily theory

The Solution: Topical Affinity Propagation

Define a function to quantify the similarity between neighborhood users

Estimate how a user can represent his neighbors

The topic information can be obtained by any tagging system or topic modeling approach

The problem is cast as identifying which node has the highest probability to influence another node on a specific topic along with the edge.
The learning task is to find a configuration for all \( \{y_i\} \) to maximize the joint probability.

**Topical Factor Graph (TFG)**

Objective function:

\[
P(v, Y) = \frac{1}{Z} \prod_{k=1}^{N} \prod_{z=1}^{T} h(y_1, \ldots, y_N, k, z) \prod_{i=1}^{N} \prod_{z=1}^{T} g(v_i, y_i, z) \prod_{e_{kl} \in E} \prod_{z=1}^{T} f(y_k, y_l, z)
\]

1. How to define?
2. How to optimize?
How to define (topical) feature functions?

- **Node feature function**

  \[ g(v_i, y_i, z) = \begin{cases} 
  \frac{w_{iz}^i y_i^z}{\sum_{j \in NB(i)} (w_{ij}^z + w_{ji}^z)} & y_i^z \neq i \\
  \frac{\sum_{j \in NB(i)} w_{ji}^z}{\sum_{j \in NB(i)} (w_{ij}^z + w_{ji}^z)} & y_i^z = i 
  \end{cases} \]

- **Edge feature function**

  \[ f(y_i, y_j) = \begin{cases} 
  w[v_i \sim v_j] & y_i = y_j \\
  1 - w[v_i \sim v_j] & y_i \neq y_j 
  \end{cases} \]

  or simply binary

- **Global feature function**

  \[ h(y_1, \ldots, y_N, k, z) = \begin{cases} 
  0 & \text{if } y_k^z = k \text{ and } y_i^z \neq k \text{ for all } i \neq k \\
  1 & \text{otherwise.} 
  \end{cases} \]
Model Learning Algorithm

\[
\begin{align*}
\text{Sum-product:} & \quad m_{y \rightarrow f}(y, z) = \prod_{f' \sim y \setminus f} m_{f' ightarrow y}(y, z) \prod_{z' \neq z} \prod_{f' \sim y \setminus f} m_{f' ightarrow y}(y, z')(\tau_{z' z}) \\
& \quad = \sum_{\sim \{y\}} \left( f(Y, z) \prod_{y' \sim f \setminus y} m_{y' \rightarrow f}(y', z) \right) \\
& \quad + \sum_{z' \neq z} \tau_{z' z} \sum_{\sim \{y\}} \left( f(Y, z') \prod_{y' \sim f \setminus y} m_{y' \rightarrow f}(y', z') \right)
\end{align*}
\]

- Low efficiency!
- Not easy for distributed learning!
New TAP Learning Algorithm

1. Introduce two new variables $r$ and $a$, to replace the original message $m$.

2. Design new update rules:

$$r_{ij}^z = b_{ij}^z - \max_{k \in NB(j)} \{b_{ik}^z + a_{ik}^z\}$$

$$a_{jj}^z = \max_{k \in NB(j)} \min\{r_{kj}^z, 0\}$$

$$a_{ij}^z = \min(\max\{r_{ij}^z, 0\}, -\min\{r_{jj}^z, 0\}) - \max_{k \in NB(j) \setminus \{i\}} \min\{r_{kj}^z, 0\}, i \in NB(j)$$

The TAP Learning Algorithm

Input: $G = (V, E)$ and topic distributions $\{\theta_v\}_{v \in V}$
Output: topic-level social influence graphs $\{G_z = (V_z, E_z)\}_{z=1}^T$

1.1 Calculate the node feature function $g(v_i, y_i, z)$;
1.2 Calculate $b_{ij}^z$ according to Eq. 8;
1.3 Initialize all $\{r_{ij}^z\} \leftarrow 0$;
1.4 repeat
   1.5 foreach edge-topic pair $(e_{ij}, z)$ do
      1.6      Update $r_{ij}^z$ according to Eq. 5;
   1.7 end
   1.8 foreach node-topic pair $(v_j, z)$ do
      1.9      Update $a_{jj}^z$ according to Eq. 6;
   1.10 end
   1.11 foreach edge-topic pair $(e_{ij}, z)$ do
      1.12      Update $a_{ij}^z$ according to Eq. 7;
   1.13 end
1.14 until convergence;
1.15 foreach node $v_t$ do
   1.16      foreach neighboring node $s \in NB(t) \cup \{t\}$ do
   1.17      Compute $\mu_{st}^z$ according to Eq. 9;
   1.18 end
1.19 end
1.20 Generate $G_z = (V_z, E_z)$ for every topic $z$ according to $\{\mu_{st}^z\}$;
Experiments

• Data&Codes: [http://arnetminer.org/lab-datasets/soinf/](http://arnetminer.org/lab-datasets/soinf/)

<table>
<thead>
<tr>
<th>Data set</th>
<th>#Nodes</th>
<th>#Edges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coauthor</td>
<td>640,134</td>
<td>1,554,643</td>
</tr>
<tr>
<td>Citation</td>
<td>2,329,760</td>
<td>12,710,347</td>
</tr>
<tr>
<td>Film (Wikipedia)</td>
<td>18,518 films 7,211 directors 10,128 actors 9,784 writers</td>
<td>142,426</td>
</tr>
</tbody>
</table>

• Evaluation measures
  – CPU time
  – Case study
  – Application
Social Influence Sub-graph on “Data mining”

<table>
<thead>
<tr>
<th>Year</th>
<th>Pairwise</th>
<th>Influence</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>Influence on Dr. Pei</td>
<td>Jiawei Han (0.4961)</td>
</tr>
<tr>
<td>2001</td>
<td>Influenced by Dr. Pei</td>
<td>Jiawei Han (0.0082)</td>
</tr>
<tr>
<td>2002</td>
<td>Influence on Dr. Pei</td>
<td>Jiawei Han (0.4045), Ke Wang (0.0418), Jianyong Wang (0.019), Xifeng Yan (0.007), Shiwei Tang (0.0052)</td>
</tr>
<tr>
<td>2003</td>
<td>Influenced by Dr. Pei</td>
<td>Shiwei Tang (0.436), Hasan M. Jamil (0.4289), Xifeng Yan (0.2192), Jianyong Wang (0.1667), Ke Wang (0.0687)</td>
</tr>
<tr>
<td>2004</td>
<td>Influence on Dr. Pei</td>
<td>Jiawei Han (0.2364), Ke Wang (0.0328), Wei Wang (0.0294), Jianyong Wang (0.0248), Philip S. Yu (0.0156)</td>
</tr>
<tr>
<td>2005</td>
<td>Influenced by Dr. Pei</td>
<td>Chun Tang (0.5929), Shiwei Tang (0.5426), Hasan M. Jamil (0.3318), Jianyong Wang (0.1609), Xifeng Yan (0.1458), Yan Huang (0.1054)</td>
</tr>
<tr>
<td>2006</td>
<td>Influence on Dr. Pei</td>
<td>Jiawei Han (0.1201), Ke Wang (0.0351), Wei Wang (0.0226), Jianyong Wang (0.018), Ada Wai-Chee Fu (0.0125)</td>
</tr>
<tr>
<td>2007</td>
<td>Influenced by Dr. Pei</td>
<td>Chun Tang (0.6095), Shiwei Tang (0.6067), Byung-Won On (0.4599), Hasan M. Jamil (0.3433), Jaewoo Kang (0.3386)</td>
</tr>
<tr>
<td>2008</td>
<td>Influence on Dr. Pei</td>
<td>Jiawei Han (0.2202), Ke Wang (0.0234), Ada Wai-Chee Fu (0.0208), Wei Wang (0.011), Jianyong Wang (0.0095)</td>
</tr>
<tr>
<td>2009</td>
<td>Influenced by Dr. Pei</td>
<td>ZhaoHui Tang (0.654), Chun Tang (0.6494), Shiwei Tang (0.5923), Zhengzheng Xing (0.5549), Hasan M. Jamil (0.3333), Jaewoo Kang (0.3057)</td>
</tr>
</tbody>
</table>

On “Data Mining” in 2009
Now, let us think…

• What are the fundamental factors behind
  – What is behind the network of social users?
  – What is behind the sentiment of social users?

What drives users’ sentiments?
Sentiment vs. Emotion

Emotion is the driving force of user’s sentiments...

Charles Darwin:
– Emotion serves as a purpose for humans in aiding their survival during the evolution.[1]

Emotion stimulates the mind 3000 times quicker than rational thought!

Potential Directions

• From sentiment to **emotion analysis**?

• Add **social theories** into emotion analysis?

• Sentiment/emotion analysis for “**Social Good**”?
Was Anna Happy When She Published This Photo On Flickr?

Anna: a girl who just graduated

A lovely doorplate

Anna

Follow
Was Anna Happy When She Published This Photo On Flickr?

It is just too sad ...

Don't be upset, you four will meet again!

Don't be upset, you four will meet again!

We have said goodbye too many times in these two days... once again, good bye our 614!
Problem

Emotion Learning Method

It is just too sad ...

don't be upset. you four will meet again!

will never forget you guys lol

we have said goodbye too many times in these two days... once again,

good bye our 614!

Comment Generation

e \sim \text{Mult}(\theta_m)

x \sim N(\mu_e, \delta_e)

Influence Generation

c \sim \text{Mult}(\lambda_d)

Image Generation

z \sim \text{Mult}(\varphi_d)

w \sim \text{Mult}(\phi_d)

Flickr Data

• 354,192 images posted by 4,807 users
  – For each image, we also collect its tags and all comments.
  – Thus we get 557,177 comments posted by 6,735 users in total

• Infer emotion of users by considering both image and tag/comments
**Emotion Inference**

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Method</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-score</th>
<th>Emotion</th>
<th>Method</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happiness</td>
<td>SVM</td>
<td>0.242</td>
<td>0.279</td>
<td>0.259</td>
<td></td>
<td>SVM</td>
<td>0.192</td>
<td>0.236</td>
<td>0.212</td>
</tr>
<tr>
<td></td>
<td>PFG</td>
<td>0.337</td>
<td>0.312</td>
<td>0.324</td>
<td></td>
<td>PFG</td>
<td>0.309</td>
<td>0.374</td>
<td>0.339</td>
</tr>
<tr>
<td></td>
<td>LDA+SVM</td>
<td>0.333</td>
<td>0.727</td>
<td>0.457</td>
<td></td>
<td>LDA+SVM</td>
<td>0.223</td>
<td>0.223</td>
<td>0.223</td>
</tr>
<tr>
<td></td>
<td>EL+SVM</td>
<td>0.367</td>
<td>0.410</td>
<td>0.388</td>
<td></td>
<td>EL+SVM</td>
<td>0.331</td>
<td>0.432</td>
<td>0.374</td>
</tr>
<tr>
<td>Surprise</td>
<td>SVM</td>
<td>0.197</td>
<td>0.037</td>
<td>0.063</td>
<td></td>
<td>SVM</td>
<td>0.204</td>
<td>0.264</td>
<td>0.230</td>
</tr>
<tr>
<td></td>
<td>PFG</td>
<td>0.349</td>
<td>0.340</td>
<td>0.345</td>
<td></td>
<td>PFG</td>
<td>0.301</td>
<td>0.408</td>
<td>0.347</td>
</tr>
<tr>
<td></td>
<td>LDA+SVM</td>
<td>0.218</td>
<td>0.048</td>
<td>0.078</td>
<td></td>
<td>LDA+SVM</td>
<td>0.211</td>
<td>0.225</td>
<td>0.217</td>
</tr>
<tr>
<td></td>
<td>EL+SVM</td>
<td>0.425</td>
<td>0.516</td>
<td>0.466</td>
<td></td>
<td>EL+SVM</td>
<td>0.371</td>
<td>0.343</td>
<td>0.356</td>
</tr>
<tr>
<td>Anger</td>
<td>SVM</td>
<td>0.188</td>
<td>0.105</td>
<td>0.135</td>
<td></td>
<td>SVM</td>
<td>0.225</td>
<td>0.365</td>
<td>0.278</td>
</tr>
<tr>
<td></td>
<td>PFG</td>
<td>0.191</td>
<td>0.142</td>
<td>0.163</td>
<td></td>
<td>PFG</td>
<td>0.357</td>
<td>0.286</td>
<td>0.317</td>
</tr>
<tr>
<td></td>
<td>LDA+SVM</td>
<td>0.222</td>
<td>0.109</td>
<td>0.146</td>
<td></td>
<td>LDA+SVM</td>
<td>0.257</td>
<td>0.278</td>
<td>0.267</td>
</tr>
<tr>
<td></td>
<td>EL+SVM</td>
<td>0.390</td>
<td>0.370</td>
<td>0.380</td>
<td></td>
<td>EL+SVM</td>
<td>0.561</td>
<td>0.617</td>
<td>0.588</td>
</tr>
</tbody>
</table>

**SVM**: regards the visual features of images as inputs and uses a SVM as a classifier.

**PFG**: considers both color features and social correlations among images.

**LDA+SVM**: first uses LDA to extract latent topics from comments, then uses visual features, topic distributions, and social ties as features to train a SVM.

*Table 2: Performance of emotion inference.*

*Averagely +37.4% in terms of F1*
To What Extend Your Friends Can Disclose Your Emotions?

- **Comments** stands for the proposed method ignoring comment information.
- **Tie** ignores social tie information.

Fear images have similar visual features with Sadness and Anger.

Homophily suggests that friends with similar interests tend to have similar understanding of disgust.

![Graph showing emotion classification](image-url)
Our model demonstrates how visual features distribute over different emotions. (e.g., images representing Happiness have high saturation)

Positive emotions attract more response (\(+4.4\) times) and more easily to influence others compared with negative emotions.
Potential Directions

• From sentiment to *emotion analysis*?

• Add *social theories* into emotion analysis?

• Sentiment/emotion analysis for “*Social Good*”? 
Summary

• From text sentiment to user sentiment
• From user sentiment to network sentiment
• From sentiment analysis to emotion analysis
• From network interaction to social influence
Related Publications

Thank you!

Collaborators: Lillian Lee, Chenhao Tan (Cornell)
Jinghai Rao (Nokia) Jimeng Sun (IBM/GIT)
Ming Zhou, Long Jiang (Microsoft)
Yuan Zhang, Jia Jia, Yang Yang, Boya Wu, Xiaohui Wang (THU)

Jie Tang, KEG, Tsinghua U,
Download all data & Codes,
http://keg.cs.tsinghua.edu.cn/jietang
http://aminer.org/data
http://aminer.org/data-sna