Exploring limits to prediction in complex social systems: Predicting cascade size on Twitter

Travis Martin
Jake Hofman, Amit Sharma
Ashton Anderson, Duncan Watts
A personal introduction

University of Michigan, Computer Science
– Network science

Summer @ Microsoft Research
– Early work on hard problem
– Please ask me questions
– WWW 2016
Predicting success on Twitter?

Bakshy, Hofman, Mason, Watts (2011):
How viral will my tweet be?
“Cascades are unpredictable!”
## Incomplete history of cascade prediction

<table>
<thead>
<tr>
<th>Who</th>
<th>Predicting</th>
<th>Features</th>
<th>Metric</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>HongD 10</td>
<td>Is item retweeted?</td>
<td>Topic Models</td>
<td>F1=0.47</td>
<td>Better than baseline</td>
</tr>
<tr>
<td>JendersKN 13</td>
<td>Will item reach some size $T$?</td>
<td>Content</td>
<td>F1&gt;0.9</td>
<td>High accuracy</td>
</tr>
<tr>
<td>TanLP 14</td>
<td>Which of two does better?</td>
<td>Wording</td>
<td>Accu=65.6%</td>
<td>Computers are OK</td>
</tr>
<tr>
<td>ChengADKL 14</td>
<td>Will cascade double?</td>
<td>Temporal</td>
<td>AUC=0.88</td>
<td>Predictable</td>
</tr>
</tbody>
</table>

Lerman, Yang, Petrovic, Romero, Kupavskii, Ma, Weng, Zhao, Yu, etc
‘Predictable’ needs a definition

1. A framework for predictability
2. Explore the predictability of information cascades (Twitter) within this framework
3. Simulation results
4. Future ideas for measuring predictability
Distinguishing model error from randomness

Empirical Observation

P[Success]

Success

“Skill World”

P[Success|skill]

Success

“Luck World”

P[Success|skill]

Success
Unpredictable: imperfect prediction with perfect model

Our two approaches for information cascades:

1. (Empirical) Does prediction performance plateau with better models and data?
2. (Simulation) Is performance highly sensitive to noise?
Why Twitter

• If we can’t predict things on Twitter, can we in the real world?
  – Lots of data
  – Fully observable spread

• Information cascades
Cascade size vs degree

Mean cascade size for a typical user vs Number of followers of a user

Number of users:
- 1
- 100
- 10,000
- 1,000,000
Our task

• Predict final # retweets of tweets with urls
• Filter to 100 popular domains
• February 2015:
  - Users: 51.6M
  - Tweets: 852M
  - Retweets: 1.806B
• Features:
  - Tweet information
  - User information
• Optimize $R^2$
  - (MSE, reduction in variance)
## Random forest features

<table>
<thead>
<tr>
<th>Model</th>
<th>Tweet time</th>
<th>Domain</th>
<th>Spam score</th>
<th>Category</th>
<th>Tweet topic</th>
<th>Past url success</th>
<th>User time</th>
<th>Followers</th>
<th>Friends</th>
<th>Statuses</th>
<th>User topic</th>
<th>Past user success</th>
<th>Topic interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Basic content</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>2. Content, topic</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>3. Content, past succ.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>4. Basic user</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>5. User, topic</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>6. User, past succ.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>7. Content</td>
<td>Dataset</td>
<td>Users</td>
<td>Tweets</td>
<td>Retweets</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. All</td>
<td>All tweets</td>
<td>51.6M</td>
<td>852M</td>
<td>1.806B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Restricted tweets</td>
<td>7.2M</td>
<td>183M</td>
<td>1.299B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Prediction limit on twitter

$R^2$ vs Model number

- All tweets
- Restricted tweets
- User success only

- Past user success
- Basic user info
How can you prove a limit?

• Results robust to other ML models
  – Decision tree, linear regression
• Consistent with prior work
• Asymptote, dependency between features
• Can’t rule everything out
  – Simulation
Simulation

• SIR disease model
• Scale free network similar to Twitter
  – 7M users, $\alpha = 2.05$
  – 8B simulated cascades
• Quality: $R_0 = \text{average neighbors infected}$
  – $p(\text{infect over edge}) \times \text{mean-degree}$
• Prediction task
  – Given (possibly noisy) estimate of $R_0$ and the seed node, predict cascade size
Increasingly heterogeneous quality
Increasing noise

\[ R^2 \quad \frac{\sigma_n}{R_0} \]

- Red: \( R_0 = 0.10 \)
- Red: \( R_0 = 0.20 \)
- Orange: \( R_0 = 0.30 \)
- Yellow: \( R_0 = 0.40 \)
1. Unifying framework for skill vs luck
2. Most extensive study of Twitter
   - Apparent limit to prediction
3. Simulation shows sensitivity to noise, heterogeneity
More ideas

1. In some cases randomness averages out
   – How/why are cascades different?
2. Are there any controlled or natural experiments we can do?
3. Better measurements of prediction goodness
   – $R^2$ is sensitive to outliers
4. More features, time dependence
   – How independent are Twitter features?
5. More realistic simulation models
Thanks!

travisbm@umich.edu
arxiv.org/abs/1602.01013
travismart.com