I. Detecting social bots

Can we verify the authenticity of users who participate in discussions and produce content in social media?

Our system detects with 94% accuracy (measured by AUC) whether a user is a bot or not, by exploiting over 800 features that capture user meta-data, social contacts, diffusion networks, content, sentiment, and temporal patterns.

II. Dataset

14.8K Legitimate users
15.8K Social bot users

Lee, Kyumin, Brian David Eoff, and James Caverlee. "Seven Months with the Devils: A Long-Term Study of Content Polluters on Twitter." ICWSM. 2011.

III. Features and feature classes

<table>
<thead>
<tr>
<th>Class (#features)</th>
<th>Description and Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network (101)</td>
<td>Retweet, Mention, HT-co-occurrence networks</td>
</tr>
<tr>
<td>User (56)</td>
<td>User meta-data.</td>
</tr>
<tr>
<td>Friend (208)</td>
<td>Contacts of a given user.</td>
</tr>
<tr>
<td>Timing (16)</td>
<td>Temporal info.</td>
</tr>
<tr>
<td>Content (274)</td>
<td>Part-of-Speech (POS) tagging &amp; Meme information</td>
</tr>
<tr>
<td>Sentiment (226)</td>
<td>Features representing sentiments.</td>
</tr>
</tbody>
</table>

IV. Feature selection

V. Performance evaluation: ROC

Performance 94% AUC!

Random Forest: 0.938
AdaBoost: 0.936
Logistic Regression: 0.847
Decision Tree: 0.812

VI. Demo


http://truthy.indiana.edu/botornot

VII. User information

VIII. Content POS tags

IX. Contact languages

X. Network

XII. Temporal Signals

http://cnets.indiana.edu

DARPA Grant Award W911NF-12-1-0037 - NSF Grant Award CCF-1101743