Multi-modal Identification of State-Sponsored Propaganda on Social Media

Xiaobo Guo, Soroush Vosoughi
{xiaobo.guo.gr, soroush}@dartmouth.edu

Minds, Machines and Society Group,
Department of Computer Science, Dartmouth
Twitter: MMS_Dartmouth

December 10, 2020
Table of Contents

1 Motivation
   ■ Background
   ■ Aim

2 Methodology
   ■ Dataset
   ■ Features
   ■ Model

3 Results
   ■ Performance
   ■ Attention Analysis
Table of Contents

1 Motivation
   ■ Background
   ■ Aim

2 Methodology
   ■ Dataset
   ■ Features
   ■ Model

3 Results
   ■ Performance
   ■ Attention Analysis
Background

**Characters**

- Propaganda
  - Purpose
  - Veracity
  - Parties
- Social Media
  - Anonymity
  - Astroturfing

**Examples**

- Russian interference in the 2016 US presidential election.
- Spreading content about the Catalan Referendum.
- Amplifying messaging supportive of the Saudi government
Our Model

- Early Detection
  - Limit Influence
- Content-based
  - User Independent
- Veracity-agnostic
  - Fake news or not
Table of Contents

1 Motivation
   ■ Background
   ■ Aim

2 Methodology
   ■ Dataset
   ■ Features
   ■ Model

3 Results
   ■ Performance
   ■ Attention Analysis
### Dataset details

<table>
<thead>
<tr>
<th>Organization</th>
<th>Data Type</th>
<th># of tweets</th>
<th>Start date</th>
<th>End date</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRA</td>
<td>training</td>
<td>4,102</td>
<td>2015-04-01</td>
<td>2016-01-19</td>
</tr>
<tr>
<td></td>
<td>validation</td>
<td>642</td>
<td>2016-01-20</td>
<td>2016-01-31</td>
</tr>
<tr>
<td></td>
<td>continuous test</td>
<td>896</td>
<td>2016-02-01</td>
<td>2016-02-28</td>
</tr>
<tr>
<td></td>
<td>delay test</td>
<td>216</td>
<td>2016-10-01</td>
<td>2017-06-30</td>
</tr>
<tr>
<td>Russian</td>
<td>training</td>
<td>5,194</td>
<td>2015-04-01</td>
<td>2015-12-14</td>
</tr>
<tr>
<td></td>
<td>validation</td>
<td>742</td>
<td>2015-12-15</td>
<td>2015-12-31</td>
</tr>
<tr>
<td></td>
<td>continuous test</td>
<td>1,534</td>
<td>2016-01-01</td>
<td>2016-02-28</td>
</tr>
<tr>
<td></td>
<td>delay test</td>
<td>6,544</td>
<td>2016-10-01</td>
<td>2017-06-30</td>
</tr>
<tr>
<td>Iranian</td>
<td>training</td>
<td>5,452</td>
<td>2015-04-01</td>
<td>2015-12-14</td>
</tr>
<tr>
<td></td>
<td>validation</td>
<td>300</td>
<td>2015-12-15</td>
<td>2015-12-31</td>
</tr>
<tr>
<td></td>
<td>continuous test</td>
<td>1,620</td>
<td>2016-01-01</td>
<td>2016-02-28</td>
</tr>
<tr>
<td></td>
<td>delay test</td>
<td>6,642</td>
<td>2016-10-01</td>
<td>2017-06-30</td>
</tr>
</tbody>
</table>

**Table**: Dataset details including organizations, number of tweets and time period.
Features

Visual Features

![Figure: Original picture](image1.png)

![Figure: Structural picture](image2.png)

Textual Features

<table>
<thead>
<tr>
<th>Text Type</th>
<th>Textual Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>#Putin’s 1st New Year’s ”achievement” in #Syria URL</td>
</tr>
<tr>
<td>Tag</td>
<td>TAG 1st New Year’s ”achievement” in TAG URL</td>
</tr>
<tr>
<td>Miss</td>
<td>1st New Year's ”achievement” in URL</td>
</tr>
<tr>
<td>Structure</td>
<td>T W W W W W W T U</td>
</tr>
</tbody>
</table>

Table: Example of four different modifications of textual content.
The structure of our model

Figure: Structure of multi-model network
Table of Contents

1 Motivation
   - Background
   - Aim

2 Methodology
   - Dataset
   - Features
   - Model

3 Results
   - Performance
   - Attention Analysis
## F1 Score with different features

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resnet-50</td>
<td>0.714</td>
<td>0.644</td>
</tr>
<tr>
<td>Inception-v3</td>
<td>0.706</td>
<td>0.639</td>
</tr>
<tr>
<td>VGG-19</td>
<td>0.681</td>
<td>0.497</td>
</tr>
<tr>
<td>Style</td>
<td>0.477</td>
<td>0.424</td>
</tr>
<tr>
<td>Content</td>
<td>0.684</td>
<td>0.519</td>
</tr>
<tr>
<td>Texture-Content</td>
<td>0.664</td>
<td>0.510</td>
</tr>
<tr>
<td>Image-Structure</td>
<td>0.604</td>
<td>0.446</td>
</tr>
</tbody>
</table>

*Table: Visual features*

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>0.854</td>
<td>0.643</td>
</tr>
<tr>
<td>Tag</td>
<td>0.803</td>
<td>0.649</td>
</tr>
<tr>
<td>Miss</td>
<td>0.788</td>
<td>0.614</td>
</tr>
<tr>
<td>Structure</td>
<td>0.715</td>
<td>0.548</td>
</tr>
</tbody>
</table>

*Table: Textual features*

*C means test on continuous test data.
*D means test on delay test data.*
ROC and AUC with different models

**Figure: IRA (C)**

**Figure: Russian (C)**

**Figure: Iranian (C)**

**Figure: IRA (D)**

**Figure: Russian (D)**

**Figure: Iranian (D)**

*C means test on continuous test data.

*D means test on delay test data.
Attention Analysis

Figure: True Positive

Figure: True Negative

Figure: False Positive

Figure: False Negative
Thank You!