Rumor Detection with Hierarchical Social Attention Network
Han Guo, Juan Cao, Yazi Zhang, Junbo Guo and Jintao Li
ACM International Conference on Information and Knowledge Management (CIKM) 2018
Buğra Öztürk
Ege Yosunkaya
Outline

- Introduction
- Related Work
- Methodology
- Experimental Results
- Conclusion
Introduction

● Rumor definition:
  ○ A rumor is defined as a story or statement in general circulation without confirmation or certainty to facts

● Dealing with massive real-time rumor post:
  ○ Investigating rumors are quite time-consuming
  ○ Automatic rumor detection are very useful
  ○ Detect early as possible

● Proposed Method:
  ○ Hierarchical network with social attention for rumor detection(HSA-BLSTM)
Related Work

- **Rumor Detection**
  - User, Message, Topic, Propagation based hand-crafted features
  - Images included in the posts
  - Conflicting viewpoints - Credibility propagation network
  - Deep Neural Networks

- **Attention Mechanism**
  - Captures importance
  - Connecting two modules with weighting

\[
\alpha_{x,y} = F(m_x, m_y) = \frac{\exp(f(m_x, m_y))}{\sum_{y} \exp(f(m_x, m_y))}
\]
Methodology

● Formulation
  ○ Classification of a group of posts that corresponds to an event
    ■ Dividing event $e$ to subevents $e = [u_1, u_2, \ldots, u_m]$ with respect to time intervals.
    ■ Each subevent, e.g. $u_i$ consists of posts $u_i = [p_{i,1}, p_{i,2}, \ldots p_{i,n}]$
    ■ Each post consists of words $p_{i,j} = [w_{i,j}^1, w_{i,j}^2, \ldots w_{i,j}^k]$
  ○ In addition to the hierarchical textual features, 22 social features

● Social Features
  ○ User Profile: Users’ credibility, reliability and reputation
  ○ Propagation: Related to the propagation pattern (e.g. average number of comments/reposts)
  ○ Post texts: Characteristics of posts including their sentiment analysis
Methodology

Bi-LSTM Network Architecture
Methodology

Bi-LSTM Network Architecture - Word Level

- Embedded word vector $w_{i,j}^k \in \mathbb{R}^d$
- Attention mechanism for words, $F_w$ score function in word level corresponds to the significance of each word
- Output of the hidden layer is multiplied with the normalized result of attention score function sum over for all connected hidden states
Methodology

Bi-LSTM Network Architecture - Post Level

- \( p_{i,j} \) encoded by Bi-LSTM to obtain hidden states
- Social feature vector used in the attention mechanism
- Output of the social attention corresponds to the importance of the post
- Output of the hidden layer is multiplied with the normalized result of attention score function and sum over for all connected hidden states
Methodology

Bi-LSTM Network Architecture - Subevent Level

- $\mathbf{u}_i$ fed to Bi-LSTM to obtain hidden states
- Social feature vector used in the attention mechanism
- Output of the social attention corresponds to the importance of the subevent
- Output of the hidden layer is multiplied with the normalized result of attention score function and sum over for all connected hidden states
Methodology

Network Architecture - Classifier

- High level representation of event vector $e$
- Fully connected layer
- Softmax function to map event vector to class (rumor, non-rumor) probabilities.
Experimental Results

● **Datasets**
  ○ Twitter
  ○ Weibo

● **Experimental Setup for Comparison**
  ○ DTC (Decision Tree Classifier)
  ○ SVM-TS
  ○ ML-GRU (Multilayer Generic GRU)
  ○ CallAtRumor
  ○ HSA-BLSTM (Proposed Model)

● **Evaluation Metrics**
  ○ Accuracy
  ○ Precision, Recall and F1 scores (Rumor and Non-rumor)
Datasets

- **Twitter**: All rumor events are crawled from Twitter by searching keywords extracted from fake news on Snopes. Part of non-rumor events are also from Snopes, and others are from two public datasets.
- **Weibo**: This dataset includes 2,313 rumors events and 2,351 non-rumors events. The rumors events are verified by Sina community management center, and the non-rumor events are gathered by crawling posts in general threads. In our experiments, we split each dataset into training set (80%) and testing set (20%). Table 2 summarizes the statistics of these two datasets.
## Datasets

<table>
<thead>
<tr>
<th></th>
<th>Weibo</th>
<th>Twitter</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Users</strong></td>
<td>1,422,140</td>
<td>125,602</td>
</tr>
<tr>
<td><strong>Posts</strong></td>
<td>1,803,891</td>
<td>245,799</td>
</tr>
<tr>
<td><strong>Rumor Events</strong></td>
<td>2,313</td>
<td>498</td>
</tr>
<tr>
<td><strong>Non-Rumors Events</strong></td>
<td>2,351</td>
<td>493</td>
</tr>
</tbody>
</table>
Experimental Setup

- **DTC** uses Decision Tree Classifier to predict the credibility on twitter. We implement this method with features based on the statistics of posts.
- **SVM-TS** utilizes a linear SVM to classify rumors on twitter and uses time-series structure to model the social feature variations.
- **ML-GRU** uses a multilayer generic GRU network to model the microblog event as a variable-length time series, which is effective for early detection of rumors.
- **CallAtRumor** presents an LSTM model to automatically identify rumors. By using the standard attention mechanism at word level, this method could detect rumors effectively.
- **HSA-BLSTM** denotes our proposed hierarchical model with social attention.
Performance Comparison

- Rank detected rumors by the predicted scores
- Obtain the weight parameters

<table>
<thead>
<tr>
<th>source post</th>
<th>related posts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pope Francis used to be a metal head. Here's a picture of him in a Black Sabbath t-shirt.</td>
<td>- Oh well, this isn't true. #bSigns Sabbath...</td>
</tr>
<tr>
<td></td>
<td>- nm, this is fake...</td>
</tr>
<tr>
<td></td>
<td>- <em>BUZZ</em> False!</td>
</tr>
<tr>
<td></td>
<td>- ...remember the pic of a younger pope francis you rt'd? With a black sabbath shirt?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>source post</th>
<th>related posts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shandong and Jiangsu is alarming! Please spread it! Do not beat the rove beetle when it is on your body...</td>
<td>- Is it true or a rumor?</td>
</tr>
<tr>
<td></td>
<td>- Is it true? Please verify and debunk the rumor...</td>
</tr>
<tr>
<td></td>
<td>- Surprise! Please verify it, it is horrible!</td>
</tr>
<tr>
<td></td>
<td>- It is coming back again, please take care!</td>
</tr>
</tbody>
</table>

Figure 3: Visualization of attention weights on example rumors (Top: from Twitter, bottom: from Weibo). Word-level weights are shown in different colors, and posts are listed according to their weights in descending order. [This figure is best viewed in color.]
Experimental Results cont.
Experimental Results cont.

Rumor Weibo

- Precision
- Recall
- F1

Legend:
- DTC
- SCM-TS
- ML-GRU
- CallAtRumor
- HSA-BLSTM
Experimental Results cont.

Rumor Twitter

- Precision
- Recall
- F1

- DTC
- SCM-TS
- ML-GRU
- CallAtRumor
- HSA-BLSTM
Experimental Results cont.

Non-Rumor Weibo

- Precision
- Recall
- F1

Legend:
- DTC
- SCM-TS
- ML-GRU
- CallAtRumor
- HSA-BLSTM
Experimental Results cont.

Non-Rumor Twitter

- DTC
- SCM-TS
- ML-GRU
- CallAtRumor
- HSA-BLSTM
Ablation Study

• **H-BLSTM**: We remove the attention mechanism and the social features in each semantic level.

• **HA-network**: The Bi-LSTM layer and the social features in each semantic level are removed.

• **HA-BLSTM**: Social features are removed.

• **HUA-BLSTM**: Only the user-profile features are used.

• **HDA-BLSTM**: Only the propagation features are used.

• **HPA-BLSTM**: Only the post-texts features are adopted.

• **HA-BLSM+S**: We simply concatenate the social features with the event representation in this model.
Ablation Study cont.
Ablation Study cont.

Rumor Weibo

- Precision
- Recall
- F1

Graph showing performance metrics for different models.
Ablation Study cont.

Rumor Twitter

- Precision
- Recall
- F1

Legend:
- H-BLSTM
- HA-Network
- HA-BLSTM
- HUA-BLSTM
- HDA-BLSTM
- HPA-BLSTM
- HS-BLSTM+S
Ablation Study cont.

Non-Rumor Weibo

- Precision
- Recall
- F1

Models:
- H-BLSTM
- HA-Network
- HA-BLSTM
- HUA-BLSTM
- HDA-BLSTM
- HPA-BLSTM
- HS-BLSTM+S
Ablation Study cont.

Non-Rumor Tweet

- Precision
- Recall
- F1

Legend:
- H-BLSTM
- HA-Network
- HA-BLSTM
- HUA-BLSTM
- HDA-BLSTM
- HPA-BLSTM
- HS-BLSTM+S
Figure 4: Early detection on Weibo dataset
Figure 5: Early detection on Twitter dataset
Conclusion

- Hierarchical Bi-LSTM network with social features as attention mechanism
- Outperforms other state-of-the-arts approaches such as ML-GRU and CallAtRumor
- Can effectively and stably detect rumor events based on a small quantity of posts, important for early detection