EMET: EMBEDDINGS FROM MULTILINGUAL-ENCODER TRANSFORMER FOR FAKE NEWS DETECTION

ICASSP May 04–08, 2020 Barcelona
Stephane Schwarz, Antônio Theóphilo, and Anderson Rocha
Institute of Computing, Unicamp, BR
Motivation

% of U.S. adults who use at least one social media site

- 2019: 72%
- 2013: 50%
- 2009: 36%
- 2005: 5%

Social media manipulation campaigns

- 2019: 70
- 2018: 48
- 2017: 28

Pew Research Center [1]

Oxford Internet Institute [2]
Motivation

Pew Research Center [1]

% of U.S. adults who use at least one social media site

Social media manipulation campaigns

Oxford Internet Institute [2]
EMET: EMBEDDINGS FROM MULTILINGUAL-ENCODER TRANSFORMER FOR FAKE NEWS DETECTION

ICASSP May 04–08, 2020 Barcelona
Stephane Schwarz, Antônio Theóphilo, and Anderson Rocha
Institute of Computing, Unicamp, BR
What is fake news?

Gene Park @GenePark
This guy keeps reusing the same pic of Obama happening now.

It's fake, it's from Thanksgiving 2015, he's in
@twitter.com/Feisal_Hagi/st...

Gene Park @GenePark
Look how desperate @feisal_hagi is attaching
same fake news tweet and RTing himself, pinn
gross. @pic.twitter.com/pW5YB4xtE

2:23 PM - Aug 27, 2017

Gene Park @GenePark
Obama is seriously not your President.

White House.

Obama is seriously not your President.

THIS IS WHAT LIE

Obama has no clue!

120 Retweets 276 Retweets

Jason Michael @Jaggit
Believe it or not, this is a shark on the freeway in Houston,
Texas. #HurricaneHarvey

3:00 AM - Aug 28, 2017

146K 93.5K people are talking about this
Prior Art: Heuristic-based

Casilillo et al. [3], Theóphilo et al. [5]
Prior Art: Multi-domain

Qi et al. [4]
EMET: Our hypothesis

Jesus words: “Then you will know the truth, and the truth will set you free.”

John 8:32
EMET: Problem statement

\[ F : \left( T_i^E \cdot C(T_i^E) \cdot N^E \right) \rightarrow y \]
EMET: Problem statement

\[ \mathcal{F} : (T_i^E \cdot C(T_i^E) \cdot N^E) \rightarrow y \]
EMET: Problem statement

Jason Michael
@Jeggit

Believe it or not, this is a shark on the freeway in Houston, Texas. #HurricaneHarvey
3:00 AM - Aug 28, 2017

146K 93.5K people are talking about this

\[ C(T_i^E) \]

\[ \mathcal{F}: (T_i^E \cdot C(T_i^E) \cdot N^E) \rightarrow y \]
$$\mathcal{F} : (T_i^E \cdot C(T_i^E) \cdot N^E) \rightarrow y$$
EMET: Problem statement

\[ \mathcal{F} : (T_i^E \cdot C(T_i^E) \cdot N^E) \rightarrow y \]
Training News was obtained from BBC and for Test set from Reuters.
## EMET: Dataset

<table>
<thead>
<tr>
<th>Class</th>
<th>Train</th>
<th>Test</th>
<th>Augmen. Train</th>
<th>Augmen. test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real</td>
<td>4314</td>
<td>2200</td>
<td>9304</td>
<td>22000</td>
</tr>
<tr>
<td>Fake</td>
<td>6690</td>
<td>3732</td>
<td>23026</td>
<td>36262</td>
</tr>
<tr>
<td>Unknown</td>
<td>1416</td>
<td>600</td>
<td>2361</td>
<td>600</td>
</tr>
</tbody>
</table>
## EMET: Dataset

<table>
<thead>
<tr>
<th>Class</th>
<th>Train</th>
<th>Test</th>
<th>Augmen. Train</th>
<th>Augmen. test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real</td>
<td>4314</td>
<td>2200</td>
<td>9304</td>
<td>22000</td>
</tr>
<tr>
<td>Fake</td>
<td>6690</td>
<td>3732</td>
<td>23026</td>
<td>36262</td>
</tr>
<tr>
<td>Unknown</td>
<td>1416</td>
<td>600</td>
<td>2361</td>
<td>600</td>
</tr>
</tbody>
</table>
EMET: Problem statement
EMET: Input
EMET: Convolution
EMET: Fully-connected
EMET: Classification Ensemble – Training
EMET: Classification Ensemble - Testing
Experimental Results: Questions

- How text embeddings from a multilingual encoder and the usage of news pieces improve the identification of misleading content on social media?

- How the comments contribute to improving classification performance?

- How the ensemble method capture general information to better model the test set?
### Experimental Setup: Unchecked news (UNC)

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCN</td>
<td>76.4</td>
<td>76.9</td>
<td>76.4</td>
<td>75.2</td>
</tr>
<tr>
<td>EANN-Text [6]</td>
<td>53.2</td>
<td>59.8</td>
<td>54.1</td>
<td>56.8</td>
</tr>
<tr>
<td>MVAE-Text [7]</td>
<td>52.6</td>
<td>52.7</td>
<td>53.9</td>
<td>53.2</td>
</tr>
</tbody>
</table>

![Diagram of the experiment setup]
Experimental Setup: Checked news (CN)

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>CN</td>
<td>92.92</td>
<td>92.99</td>
<td>92.92</td>
<td>92.94</td>
</tr>
<tr>
<td>EANN-Text</td>
<td>53.2</td>
<td>59.8</td>
<td>54.1</td>
<td>56.8</td>
</tr>
<tr>
<td>MVAE-Text</td>
<td>52.6</td>
<td>52.7</td>
<td>53.9</td>
<td>53.2</td>
</tr>
</tbody>
</table>
Experimental Results: Questions

- How text embeddings from a multilingual encoder and the usage of news pieces improve the identification of misleading content on social media?

- How the comments contribute to improving classification performance?

- How the ensemble method capture general information to better model the test set?
Experimental Setup: **CN and comments (CNC)**

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>CN</td>
<td>92.92</td>
<td>92.99</td>
<td>92.92</td>
<td>92.94</td>
</tr>
<tr>
<td>CNC</td>
<td>93.47</td>
<td>93.91</td>
<td>93.47</td>
<td>93.61</td>
</tr>
</tbody>
</table>
Experimental Results: Questions

● How text embeddings from a multilingual encoder and the usage of news pieces improve the identification of misleading content on social media?

● How the comments contribute to improving classification performance?

● How the ensemble method capture general information to better model the test set?
Experimental Setup: **CNC and ensemble (CNCE)**

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNC</td>
<td>93.47</td>
<td>93.91</td>
<td>93.47</td>
<td>93.61</td>
</tr>
<tr>
<td>CNCE</td>
<td><strong>94.08</strong></td>
<td>91.31</td>
<td>91.21</td>
<td>91.26</td>
</tr>
</tbody>
</table>
Conclusion

EMET helps to address the problem of fake news detection on social media platforms in a multilingual scenario.
Explore Multi Domain data (Image + Text)
Acknowledgments
References


EMET: EMBEDDINGS FROM MULTILINGUAL-ENCODER TRANSFORMER FOR FAKE NEWS DETECTION

ICASSP May 04–08, 2020 Barcelona
Stephane Schwarz, Antônio Theóphilo, and Anderson Rocha
Institute of Computing, Unicamp, BR