NLP/UIOVA @ SemEval-2020 Task8: You’re not the only one cursed with knowledge—Multi-branch model mememotion analysis

Ingroy Shrestha, Jonathan Rusert
University of Iowa

INTRODUCTION
- With rise in social media culture, sharing of internet memes on social media platform has increased.
- Memes are used to express humor, embarrassment, hate and even more emotions.
- Hateful or offensive memes can be created and lead to increase in hate crime, which motivates for detection of offensive memes (Heikkilä, 2017; Sabat et al., 2019).
- Meme understanding requires both visual and language understanding.
- Manual monitoring of memes is not scalable using human resources.
- Memotion Analysis: Automatic classification of memes (Sharma et al., 2020).

DATA DESCRIPTION

<table>
<thead>
<tr>
<th>Subtask A</th>
<th>Subtask B</th>
</tr>
</thead>
<tbody>
<tr>
<td>sentiment</td>
<td># of instances</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>positive</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>negative</td>
<td></td>
</tr>
<tr>
<td>neutral</td>
<td></td>
</tr>
</tbody>
</table>

Subtask C
- scale of semantic ( # of instances)
  - not (0) slightly (1) mildly (2) very (3)
  - humorous: 1649 2452 2236 649
  - sarcastic: 1543 3503 1546 394
  - offensive: 2709 2591 1465 221

METHODS
- Proposed Models
  - Hybrid Model Weighted (HybridE)
  - Hybrid Weighted Average (HybridW)
- Baseline
  - Logistic regression for both image and text classification

DISCUSSION
- Tradeoff in the performance of BLSTM and CNN
  - BLSTM better performance in macro F1 and CNN better performance in μ F1.
- Comparison of HybridE and HybridW
  - HybridW > HybridE (macro F1)
- Class imbalance and effect on performance metric
  - μ F1 > macro F1
- Failure of transfer learning
  - Overfitting: complexity of pre-trained architecture or, failure to learn task specific features provided small train set.

CONCLUSION & LIMITATIONS
- Individual system: BLSTM (text) and CNN (image)
- Proposed system (hybrid approach)
  - HybridE (equal weightage to prediction probability of BLSTM and CNN)
  - HybridW (weighted average based on BLSTM and CNN performance).
- HybridE performs better on overall than individual system. HybridW shows little of no improvement over HybridE.
- Limitations
  - Trained models for text and image separately; feeding text and image output into a common dense layer; catch some features missed.
  - Explore problem on larger dataset.
  - Explore optimal configuration for transfer learning.

CONTACT INFO
- Ingroy Shrestha (ingroy-shrestha@uiowa.edu)
- Jonathan Rusert (jonathan.rusert@uiowa.edu)

REFERENCES