NLP_UIOWA @ SemEval-2020 Task8: You're not the only one cursed with knowledge –Multi branch model memotion analysis

Ingroj 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.

An interesting Squid

- Hateful or offensive memes can be created and lead to increase in hate crime, which motivates for detection of offensive memes (Heikkil¨a, 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

Subtask A	
sentiment	**# of instances**
positive	4160
negative	631
neutral	2201

Subtask B		
semantic	**present**	**not**
humorous	5337	1649
sarcastic	5443	1543
offensive	4277	2709
motivational	2465	4521

Subtask C

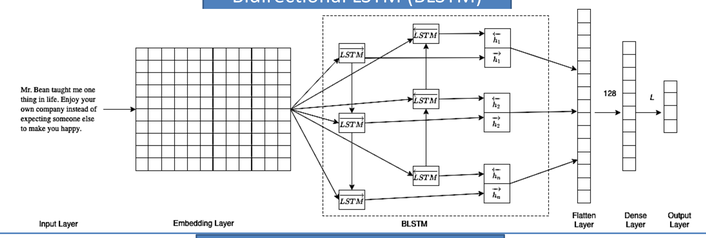
		scale of semantic (# of instances)			
		not (0)	**slightly (1)**	**mildly (2)**	**very (3)**
semantic	humorous	1649	2452	2236	649
	sarcastic	1543	3503	1546	394
	offensive	2709	2591	1465	221

METHODS

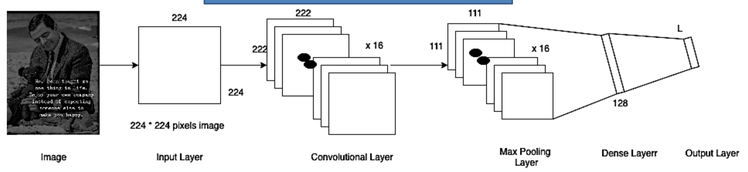
- **Proposed Models**
  - Hybrid Model Weighted (HybridE)
  - Hybrid Weighted Average (HybridW)
- **Baseline**
  - Logistic regression for both image and text classification
- **Transfer Learning**
  - BERT-base (Text)
  - VGG-16 (Image)
  - ResNet 18 (Image)

PROPOSED APPROACH

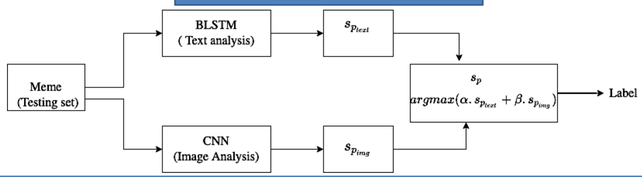
Bidirectional LSTM (BLSTM)



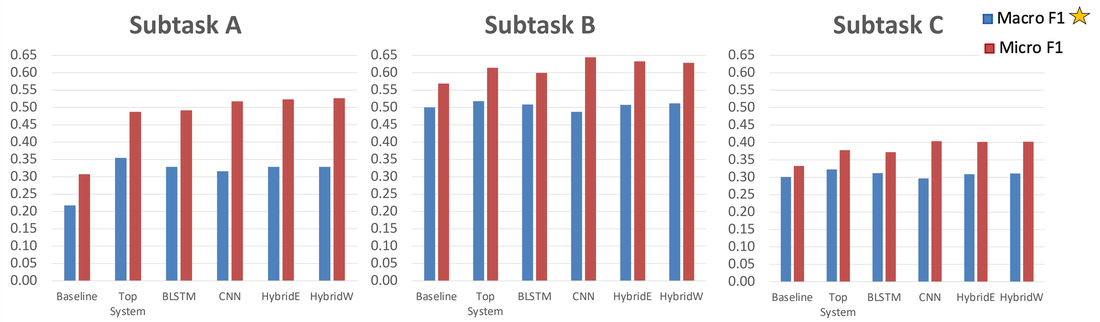
CNN



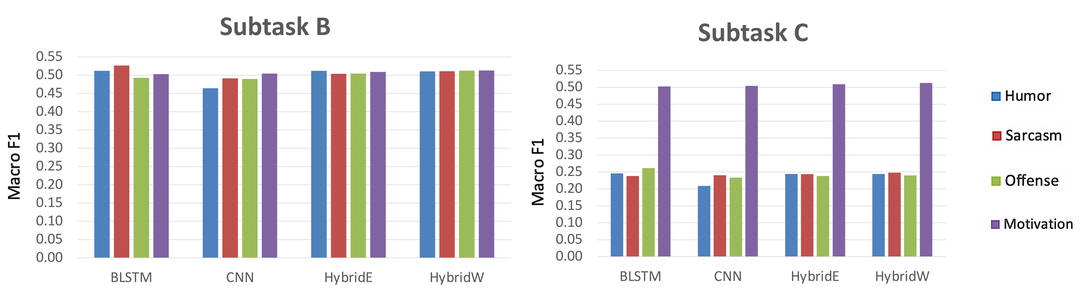
Hybrid approach



PERFORMANCE (Overall)



PERFORMANCE (Class-wise)



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 $<^{\varepsilon}$ 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

- **Ingroj Shrestha** (ingroj-shrestha@uiowa.edu)
- **Jonathan Rusert** (jonathan-rusert@uiowa.edu)

REFERENCES

- Chhavi Sharma, Deepesh Bhageria, William Paka, Scott, Srinivas P Y K L, Amitava Das, Tanmoy Chakraborty,Viswanath Pulabaigari, and Bj¨orn Gamb¨ack. 2020. SemEval-2020 Task 8: Memotion Analysis-The Visuo-Lingual Metaphor! In *Proceedings of the 14th International Workshop on Semantic Evaluation (SemEval- 2020)*, Barcelona, Spain, Sep. Association for Computational Linguistics.
- Benet Oriol Sabat, Cristian Canton Ferrer, and Xavier Giro-i Nieto. 2019. Hate Speech in Pixels: Detection of Offensive Memes towards Automatic Moderation. *arXiv preprint arXiv:1910.02334*.
- Niko Heikkil¨a. 2017. Online antagonism of the alt-right in the 2016 election. *European journal of American studies* , 12(12-2).
- Marcos Zampieri, Shervin Malmasi, Preslav Nakov, Sara Rosenthal, Noura Farra, and Ritesh Kumar. 2019.
- SemEval-2019 Task 6: Identifying and Categorizing Offensive Language in Social Media (OffensEval). In *Proceedings of The 13th International Workshop on Semantic Evaluation (SemEval)*.