“A Little Birdie Told Me ... ” - Social Media Rumor Detection

Karthik Radhakrishnan* Tushar Kanakagiri* Sharanya Chakravarthy* Vidhisha Balachandran

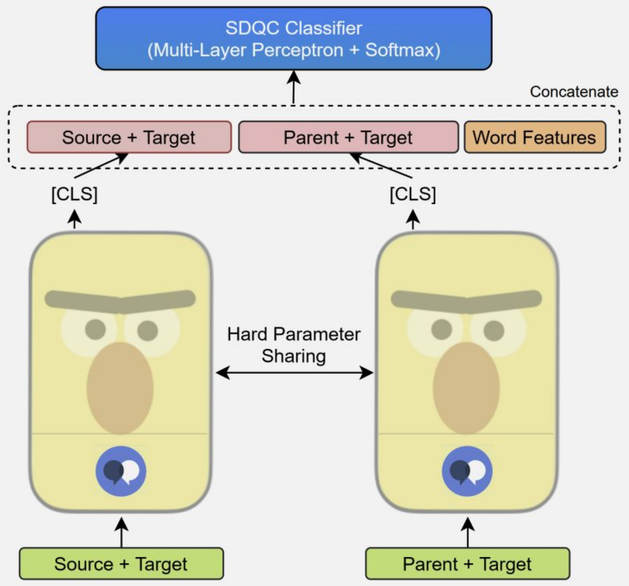
Abstract

The rise in the usage of social media has placed it in a central position for news dissemination and consumption. This greatly increases the potential for proliferation of rumours and misinformation. In an effort to mitigate the spread of rumours, we tackle the related task of identifying the stance (Support, Deny, Query, Comment) of a social media post. Unlike previous works, we impose inductive biases that capture platform specific user behavior. These biases, coupled with social media fine-tuning of BERT allow for better language understanding, thus yielding an F1 score of 58.7 on the SemEval 2019 task on rumour stance detection.

Task - Rumor Stance Classification

- **Support** the veracity of the rumor
- **Deny** the veracity of the rumor
- **Query** for additional evidence in relation to the veracity of the rumor
- **Comment** without contributing to assessing the veracity

System Description



Late Fusion of Source and Parent [CLS] representations

Conversational BERT to understand conversational constructs

Domain separation by training separate models on Twitter and Reddit data

Additional TF-IDF features for each class, for use with BERT

Transition priors from training data during post-processing

Results and Discussion

Validation Set Results

Model	Macro-F1
BLCU (Best) - Non-ensemble SOTA	56.6
Ours (Average)	56.7
Ours (Best)	**58.7**

Ablation Study

Model	Macro-F1
Base Model	51.2
+ Conversational BERT	53.7 (+2.5)
+ TF-IDF Features	55.2 (+1.5)
+ Domain Sep., Late Fusion	56.4 (+1.2)
+ Transition Priors	**58.7 (+2.3)**

Unsolvable Examples - Noisy annotations in the dataset, different labels to similar text, deleted tweets

Future Work - Utilize model for veracity detection, expand inductive biases to fact verification / conversation derailment detection

Carnegie Mellon University
Language Technologies Institute