Intermediate-Task Transfer Learning: Leveraging Sarcasm Detection for Stance Detection

Gibson Nkhata, Susan Gauch

The Sixteenth International Conference on Information, Process, and Knowledge Management

eKNOW 2024

Presenter: Gibson Nkhata

University of Arkansas, USA

gnkhata@uark.edu

Presenter Biography

- **Gibson Nkhata** received a Master’s degree in Computer Science from the University of Arkansas, USA, in 2022. He is currently a doctoral student in Computer Science at the university. He is also working as a Graduate Research Assistant in the Electrical Engineering and Computer Science (EECS) department under the DART project.

Research Interests

- Our research interest lies in the application of Deep Learning techniques and pre-trained language models in Natural Language Processing (NLP) tasks like Sentiment Analysis, Stance Detection, and rumor veracity detection on social media.

Outline

- Motivation
- Illustrative Example
- Our approach
- Experiments and Results
- Contributions and Conclusion
- Future Work

UNIVERSITY OF ARKANSAS

Motivation

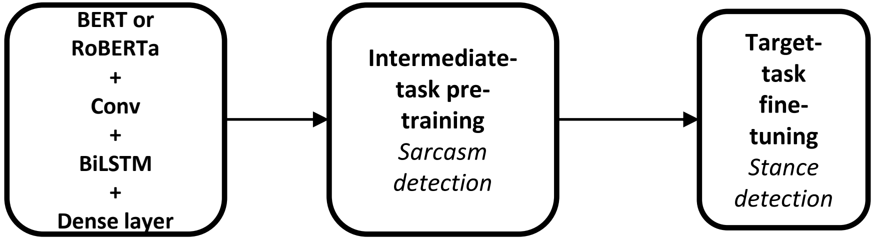
- **Stance Detection (SD)**
- SD is the automated identification of an individual's stance on a specific topic based solely on their utterance or authored material.
- Stance labels categorize expressions into *InFavor* (*Support*), *Against*, or *None*.
- Social media platforms provide a vast amount of information on diverse topics fostering SD in social, business, and political applications.
- The inclusion of sarcastic and figurative language in posts drastically impacts the performance of SD models.

Illustrative Example

- “*I like girls. They just need to know their place*”.
  - Comment based on *Feminist Movement*.
  - Sarcastic (*Against*) comment.
  - But looks positive (*Support*).
- Human evaluator can easily imply its stance but hard for a machine without prior sarcasm knowledge.
- That’s why this work proposes sarcasm detection pre-training before SD.

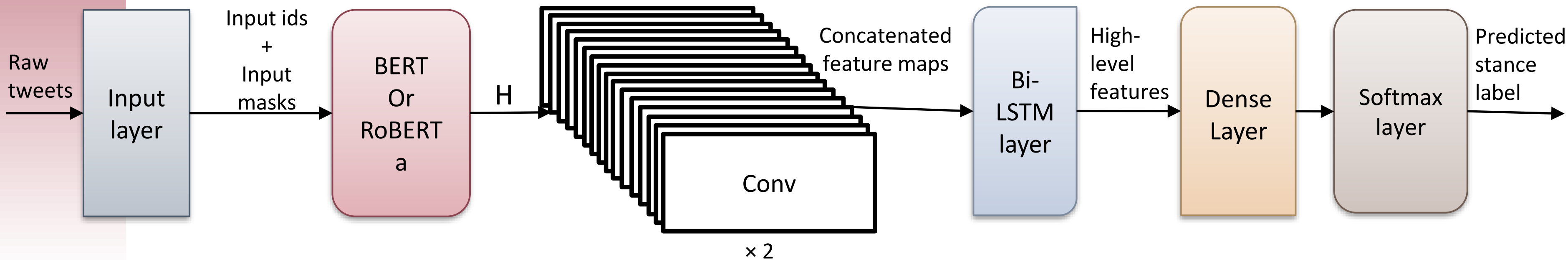
Our approach

- Intermediate-Task Transfer Learning
  - Target Task: SD
  - Intermediate Task: Sarcasm Detection
    - Sarcastic or not sarcastic



Our approach

- Underlying Model Architecture
  - Input layer
  - Embedding layer: BERT or RoBERTa
  - Deep Neural Networks: CNN, BiLSTM, and Dense layer
- Problem formulation
  - Given a tweet *T*, represented as a word sequence (*w*1,*w*2,*w*3, ...*wL*), with *L* denoting the sequence length, predict its stance regarding a given target.
  - Stance labels are categorized as *InFavor* (supporting the target/topic/claim), *Against* (opposing the topic), or *None* (indicating neutrality towards the target).



Experiments

Sarcasm Datasets

Name	Size
Sarcasm V2 Corpus (SaV2C)	3,260
The Self-Annotated Reddit Corpus (SARC)	1.1M
SARCTwitter (ST)	994

SD Datasets

Name	Size
SemEval 2016 Task 6A Dataset (SemEval)	4,063
Multi-Perspective Consumer Health Query Data (MPCHI).	1,535

Experimental Settings

Batch size	Seq length	Learning rate	epochs	optimizer	Loss function
16	24	3e-5 to 1e-9, 1e-10	10-50	Adam	Binary cross entropy loss

Results

TABLE II

EXPERIMENTAL RESULTS WITHOUT SARCASM DETECTION PRE-TRAINING

Model	SemEval						MPCHI					
	AT	**CC**	**FM**	**HC**	**LA**	**Avg**	**MMR**	**SC**	**EC**	**VC**	**HRT**	**Avg**
Sem-TAN-	0.596	0.420	0.495	0.543	0.603	0.531	0.487	0.505	0.564	0.487	0.467	0.502
Sem-CNN	0.641	0.445	0.552	0.625	0.604	0.573	0.524	0.252	0.539	0.524	0.539	0.476
Com-BiLSTM	0.567	0.423	0.508	0.533	0.546	0.515	0.527	0.522	0.471	0.474	0.469	0.493
ZSSD	0.565	0.389	0.546	0.545	0.509	0.511	-	-	-	-	-	-
Com-BERT	0.704	0.466	0.627	0.620	0.673	0.618	0.701	0.691	0.710	0.617	0.621	0.668
ChatGPT	-	-	0.690	**0.780**	0.593	0.687	-	-	-	-	-	-
Ours-RoBERTa	0.740	**0.775**	0.689	0.683	0.696	0.712	0.692	0.687	0.700	0.701	0.698	0.695
Ours-BERT	**0.767**	0.755	**0.697**	0.704	**0.702**	**0.725**	**0.747**	**0.722**	**0.704**	**0.702**	**0.732**	**0.721**

TABLE III

EXPERIMENTAL RESULTS WITH SARCASM-DETECTION PRE-TRAINING

Task	SemEval						MPCHI					
	AT	**CC**	**FM**	**HC**	**LA**	**Avg**	**MMR**	**SC**	**EC**	**VC**	**HRT**	**Avg**
SaV2C	0.595	0.718	0.596	0.645	0.578	0.626	0.605	0.545	0.545	0.352	0.495	0.508
SARC	0.697	0.612	0.683	0.557	0.641	0.638	0.605	0.545	0.545	0.352	0.495	0.508
ST	**0.769**	**0.800**	**0.774**	**0.795**	**0.741**	**0.775**	**0.749**	**0.727**	**0.704**	**0.703**	**0.739**	**0.724**

Results

TABLE IV
EXPERIMENTAL RESULTS OF AN ABLATION STUDY

Model	SemEval						MPCHI					
	AT	**CC**	**FM**	**HC**	**LA**	**Avg**	**MMR**	**SC**	**EC**	**VC**	**HRT**	**Avg**
BERT	0.674	0.677	0.678	0.609	0.685	0.665	0.568	0.519	0.441	0.482	0.595	0.521
BERT+Conv+BiLSTM	0.767	0.755	0.697	0.704	0.702	0.725	0.747	0.722	**0.704**	0.702	0.732	0.721
ST+BERT	0.712	0.735	0.698	0.687	0.696	0.706	0.687	0.601	0.540	0.466	0.546	0.568
ST+BERT+Conv	**0.770**	0.759	0.689	0.683	0.694	0.719	0.458	0.535	0.479	0.350	0.524	0.469
ST+BERT+BiLSTM	0.747	0.765	0.675	0.657	0.678	0.704	0.640	0.618	0.573	0.528	0.633	0.598
ST+BERT+Conv+BiLSTM	0.769	**0.800**	**0.774**	**0.795**	**0.741**	**0.775**	**0.749**	**0.727**	**0.704**	**0.703**	**0.739**	**0.724**

Contributions and Conclusion

- We introduced a transfer-learning framework that leverages sarcasm detection for SD.
- The model underwent separate pre-training on three sarcasm-detection tasks before fine-tuning on two target SD tasks.
- We separately fine-tuned RoBERTa and BERT and sequentially concatenated them with other deep neural networks.
- BERT models gave promising results.
- We established the correlation between sarcasm detection and SD through failure analysis.
- We showed that not every sarcasm-detection intermediate task improved SD due to incongruous linguistic attributes.
- To the best of our knowledge, this is the inaugural exploration of sarcasm-detection pre-training applied to the BERT(RoBERTa)+Conv+BiLSTM architecture before SD finetuning.

Future work

- We will assess variant BERT or RoBERTa embeddings tailored to health-related text data.
- We will also concentrate on cross-target SD.
- We will do a more comprehensive examination of other intermediate tasks, including sentiment and emotion knowledge.