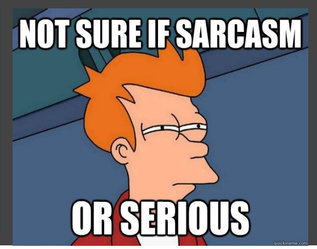

Sarcasm Detection With Context Information

Lydia Xu (leediaxu@stanford.edu), Vera Xu (veraxrl@stanford.edu)

Problem

- Sarcasm Detection is critical in understanding sentiments and intentions
- Existing approaches: SVM, various LSTMs
- Previous models detect sarcasm based on utterance in isolation. **Can we explore the role of conversational context in sarcasm detection?**
- Apply state-of-the-art models on sarcasm detection using both context and response data

Data

- Discussion Forum
  - 4692 sarcastic and non-sarcastic (balanced)
  - Tagged by crowdsourcing
- Reddit
  - 1.3 million self-annotated corpus（balanced)

Both datasets contain context texts, response texts and sarcastic label.

Approach

- Vanilla LSTM with only response information
- Attention-based LSTM with both context and response using sentence-level attention (Figure 1)
- Conditional LSTM where the initial hidden state of the response LSTM is the last hidden state of the context LSTM (Figure 2)
- BERT Classifier

$$v = \sum_{i \in [1,d]} \alpha_i h_i$$

$$\alpha_i = \frac{exp(u_i^T u_s)}{\sum_{i \in [1,d]} exp(u_i^T u_s)}$$

$$u_i = tanh(W_s h_i + b_s)$$

(Attention-based LSTM equations)

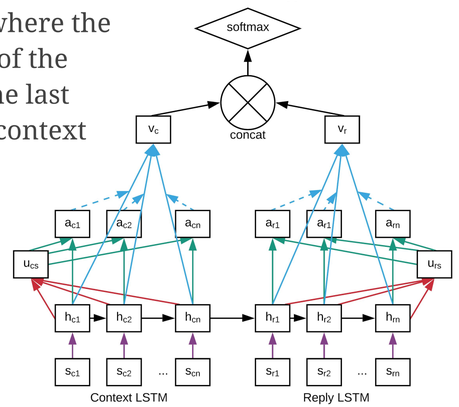

Figure 1. Attention-based LSTM Architecture

Results

- Contextual information help improves performance of sarcasm detection
- BERT outperformed LSTMs by large margin
- Conditional LSTM generates best metrics among various LSTM models

Table 1. Precision, Recall, F1 and Accuracy for all experiments

Data	Experiment	S			NS			Acc
		P	R	F1	P	R	F1	
Discussion Forum	SVM_r (Baseline)	65.55	66.67	66.1	66.1	64.96	65.52	n/a
	SVM_{c+r} (Baseline)	63.22	61.97	62.63	62.77	64.1	63.5	n/a
	$LSTM_r$	67.76	60.8	64.01	62.75	69.54	65.97	**65.05**
	$LSTM_c$ + $LSTM_r$ with attention	51.44	84.71	64.01	43.51	12.83	19.82	51.22
	$LSTM_{conditional}$	**71.32**	**80.13**	**75.47**	**75.82**	**65.93**	**70.53**	**73.23**
	$LSTM_{conditional}$ [Ghosh et al., 2017]	70.03	76.92	73.32	74.41	67.1	70.56	n/a
	$BERT_r$	68.6	84.14	75.58	44.94	25.15	32.25	64.1
	$BERT_{r+c}$	**70.71**	**87.01**	**78.02**	**55.06**	**30.62**	**39.35**	**67.74**
Reddit	**$LSTM_r$**	57.64	80.95	67.33	68.28	40.8	51.08	**60.83**
	$LSTM_c$ + $LSTM_r$ with attention	50.56	65.18	56.95	47.07	32.7	38.59	49.38
	$LSTM_{conditional}$	**67.86**	**66.58**	**67.22**	**66.83**	**68.11**	**67.47**	**67.34**
	$BERT_r$	**82.23**	**78.25**	**76.18**	**80.02**	**92.18**	**85.67**	**76.08**
	$BERT_{r+c}$	64.79	60.93	62.8	74.99	77.96	76.45	71.18

Parameter tuning: we tried multiple hidden sizes and data sizes for training the models and found out that for both cases, too small sizes lead to lack of information while too large sizes lead to overfitting.

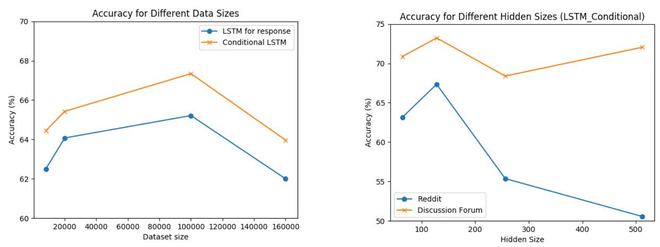

Figure 3. Parameter Tuning Experiments

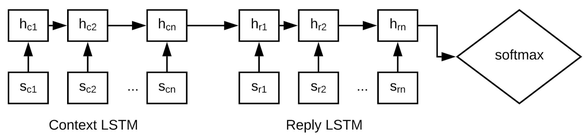

Figure 2. Conditional LSTM Architecture

Analysis

- LSTM models has greater performance gap between datasets, while BERT has similar performances. Reddit Data has a lot of internet slangs and informal languages. It is hard for our Word Embedding and LSTM model to pick up words not available in GloVe, for example, responses like "NO WAIIIIIIIIIII".
- Usually, the usage of certain markdowns and emojis gives strong indication of sarcasm in forums. However, in our pre-processing, these patterns are either being tokenized into different tokens or treated as words not found in the pre-trained embeddings. See "emoticonX-Good", "emoticonX-Confused" and ** examples in Table 2.

Table 2. Context and Response Examples with Labels

Dataset	Context	Reply
Discussion Forum	Obviously you missed the point. So sorry the the irony was beyond you.	I guess we all missed your point Justine, whatever it might have been. emoticonXConfused Better luck next time. emoticonXGood
Reddit	BLM is trying to shut down MSP International Airport	An hour ago I thought black lives *didn't* matter, but now that they've protested at the airport I've completely changed my mind and now think **black lives matter**

Conclusion

- For LSTM models, utilizing context information greatly improves model performance.
- Our conditional LSTM model establishes the new state-of-art performance.
- Pre-trained BERT classifier can achieve great results for sarcasm detection, even with small dataset. With larger dataset like Reddit data, it outperforms our best LSTM model by 7%.

Reference: Debanjan Ghosh, Alexander Richard Fabbri, and Smaranda Muresan. The role of conversation context for sarcasm detection in online interactions. SIGDial, 2017.