Character-based Neural Networks for Sentence Pair Modeling

Wuwei Lan and Wei Xu

Introduction
- Sentence pair modeling is critical for paraphrase identification, question answering, natural language inference and etc.
- Various neural models achieved state-of-the-art performance by using pretrained word embeddings, however they have poor coverage in domain (e.g., social media) with high OOV ratio.
- We explored character-based neural networks for sentence pair modeling, which is more challenging than individual sentence modeling: similarly spelled words with completely different meanings could introduce error (e.g., ware and war).

Example for Sentence Pair Modeling
Paraphrase task: given a sentence pair, predict whether they imply the same meaning. Sample from Twitter URL corpus [Lan et.al 2017]:

**Context modeling:**

\[
\hat{h}_t = \text{LSTM}(w_t, \hat{h}_{t-1})
\]

\[
\hat{h}_t = \text{LSTM}(w_t, \hat{h}_{t-1})
\]

\[
\hat{h}_t = [\hat{h}_t, \hat{h}_t]
\]

\[
\hat{h}_t = \hat{h}_t + \hat{h}_t
\]

**Pairwise word interaction:**

\[
D(\hat{h}_t, \hat{h}_j) = \cos(\hat{h}_t, \hat{h}_j), \quad \text{L2Euclid}(\hat{h}_t, \hat{h}_j), \quad \text{DotProduct}(\hat{h}_t, \hat{h}_j)
\]

**Similarity focus:**
- sorting the interaction values and selecting top ranked pairs

**Aggregation and prediction:**
- 19-layer deep ConvNet

Embedding Subwords in PWIM

1. **Char RNN** [Ling et.al 2015]

On the Mat **There** Sit Cats

2. **Char CNN** [Kim et.al 2016]

On the Mat **There** Sit Cats

Language Modeling Objective (Multitask)

The forward LSTM predicts the next word and the backward LSTM predicts the previous word. The log-likelihood loss for both language models is added to the training objective:

\[
\hat{E} = E + \gamma (E + \tilde{E})
\]

Experiments

We performed experiments on three benchmark datasets for paraphrase identification: Twitter URL (social media/news, OOV ratio 31.5%), PIT-2015 (social media, OOV ratio 13.7%) and MSRP (news, OOV ratio 9.0%).

Conclusion

✓ Pretrained word embedding is not a necessity for sentence pair modeling.
✓ Subword models without any pretraining achieved new SOTA results in Twitter URL and PIT-2015.
✓ Multitask LM can improve subword model performance by injecting semantic information.

Source Code

https://github.com/lanwuwei/Subword-PWIM

See Also

- [He et.al 2016](https://github.com/lanwuwei/Subword-PWIM)