Semi-supervised Multitask Learning for Sequence Labeling
Marek Rei
University of Cambridge

Sequence Labeling

The task:
Given a sequence of tokens, predict a label for every token.

Named Entity Recognition:

POS-tagging:

Error Detection:

+ + + - + + + + - +
I like to playing the guitar and sing louder.

Neural Sequence Labeling

• Sequence of tokens mapped to word embeddings.
• Bidirectional LSTM builds context-dependent representations for each word.
• A small feedforward layer encourages generalisation.

• Conditional Random Field (CRF) at the top outputs the most optimal label sequence for the sentence.
• Using character-based dynamic embeddings (Rei et al., 2016) to capture morphological patterns and unseen words.

Multitask Learning

• Sequence labeling datasets can be very sparse: only 17% of tokens in CoNLL-03 are a named entity.
• We want an additional objective that makes full use of the data to learn features for semantic composition.
• Language modeling 1) requires no extra annotation, 2) has a large number of possible targets for each position.

• The network predicts the next word together with the main label.
• Cannot simply add it as an extra output layer – the next word is already given as input to the nextwork.

Language Modeling Objective

• The forward-moving LSTM predicts the next word in the sequence.
• The backwards-moving LSTM predicts the previous word in the sequence.
• Both LSTMs predict the target label.

\[ \hat{E} = E + \gamma (\hat{E} + \tilde{E}) \]
\[ \hat{E} = - \sum_{t=1}^{T-1} \log(P(w_{t+1}|m_t)) \]
\[ \tilde{E} = - \sum_{t=2}^{T} \log(P(w_{t-1}|\tilde{m}_t)) \]

Analysis

• Visualising convergence on the FCE development set after each training epoch.
• LM objective improves performance at all stages of training.
• Additional parameter matrices are required for the two language models during training.
• However, the LM components are not needed during testing.
• The resulting model has the same structure and the same number of parameters as the baseline.

Conclusion

• Integrated a language modeling objective into a neural sequence labeling architecture.
• Requires no additional data and the trained model has no additional parameters.
• Provides consistent improvements on 10 different datasets.
• The source code: https://github.com/marekrei/sequence-labeler

Results

• Experiments on 10 different datasets and 4 different tasks: error detection, named entity recognition, chunking, and POS tagging.