Part 2 – Report

Name

Simon Raviv

Definitions

I will use formal language description over the English alphabet.

- Let $\Sigma 1 = \{'a', \ldots, 'z'\}$
- Let $\Sigma 2 = \{'(', ')', '[', ']', '\{', '\}'\}$
- Let the maximum sequence length to be 150 chars.

Challenging cases

Divisible without reminder

Language description

$L1 = \{a^n, b^m, c^k, d^l\}, s.t.\ n\ mod\ 2 = 0, m\ mod\ 3 = 0, k\ mod\ 5 = 0, l\ mod\ 7 = 0$

Reasons for hard to distinguish

In this language we need to be able to divide. As known, LSTM can count, but it will not be able to count divisible without reminder number of characters.

LSTM train

Parameters

- Train/dev set size: 900/100
- Epochs: 300

Generalization

The LSTM couldn't learn this language, it failed to get higher than random accuracy on both train and dev set.

Mirror sub-sequence

Language description

$L2 = \{\delta\alpha\delta^* \mid \delta\ \epsilon\ \Sigma 1, \alpha\ \epsilon\ \Sigma 1 \backslash \delta\}$

Reasons for hard to distinguish

In this language we need to be able to maintain counting state for the first half of the sequence and another counting state for the second half of the sequence. It means remembering counts from past and future directions. The regular LSTM i.e., one directional LSTM, it can't look at the sequence in both directions.

LSTM train

Parameters

- Train/dev set size: 900/100
- Epochs: 300

Generalization

The LSTM couldn't generalize on this language, it failed to get higher than random accuracy on dev set, but able to get high accuracy (over 90%) on train set, due to overfitting.

Correct parentheses phrase

Language description

$L3 = \{[[\alpha\varepsilon\alpha]^+[\alpha\beta\varepsilon\beta\alpha]^+]^* \mid \varepsilon \in \Sigma 1, \alpha, \beta \in \Sigma 2\}$

Reasons for hard to distinguish

In this language we need to be able to maintain stack like state counting, which is not possible with the regular LSTM. $L3$ is a context-free language which cannot be captured by a simple counting mechanism, a model capable of recognizing this language must contain a stack-like component.

LSTM train

Parameters

- Train/dev set size: 900/100
- Epochs: 300

Generalization

The LSTM couldn't lean this language, it failed to get higher than random accuracy on both train and dev set.