

HW4: Word Segmentation

Virgile Audi
vaudi@g.harvard.edu
Nicolas Drizard
nicolasdrizard@g.harvard.edu

March 31, 2016

1 Introduction

The goal of this assignment is to implement recurrent neural networks for a word segmentation task. The idea is to identify the spaces in sentence based on the previous characters only. This could be particularly helpful for processing languages written without spaces such as Korean or Spanish

2 Problem Description

Description of the task with example

3 Model and Algorithms

3.1 Count-based Model

3.2 Neural Language Model

3.3 Algorithm to generate spaces sequences

Greedy

Viterbi

3.4 Recurrent Neural Networks

Generic RNN Transducer

LSTM

GRU

4 Experiments

4.1 Count-based Model

4.2 Neural Language Model

4.3 Recurrent Neural Networks

4.4 Ensemble Method

5 Conclusion

End the write-up with a very short recap of the main experiments and the main results. Describe any challenges you may have faced, and what could have been improved in the model.