An Investigation into Character-Aware Neural Language Models

Anonymous

November 7, 2023

Abstract

This document explores the conceptual approach and implementation details of a Character-Aware Neural Language Model, based on the work of Kim et al. (2016) [1]. It discusses implementation details that will be completed using various machine learning models in PyTorch.

1 Introduction

Character-Aware Neural Language Models represent another advancement in allowing computers to understand subtleties at the character level in language processing tasks. Combining character-level information with traditional word embeddings, this model effectively captures spatial and and semantic nuances of language beyond the capabilities of models that only tokenize at the word-level.

2 Algorithm Discussion

The Character-Aware Neural Language Model is designed to utilize both character representations when structuring neural networks instead of words. For characters a 1D convolutional neural network (CNN) extracts spatial features. These features have to be sent through a highway network allowing for adaptive information flow.

Once processed, these features are passed to a recurrent neural network whether it be a Long-Short-Term Memory (LSTM) model or Gated Recurrent Unit (GRU) model that models language sequences.

3 Implementation Discussion

The expected neural network structure will be in the order of the following: a character embedding layer, a Char-CNN, a highway network, and an LSTM or GRU as the RNN. The embedding layer will transform character indices into vectors, using ascii characters as the one-hot vectors, which the Char-CNN will

process. The characters will come from word tokens. The principle components from the CNN will be extracted through a highway network designed to control the flow of information and build dependencies over words.

4 Challenges and Design of the Highway Model

One of the primary challenges lies in designing the highway network, which must effectively deduce character-level features from words while retaining important information regardless of the repetition of a single character.

To address the complexity of training such a network, the design will incorporate a series of transformation and carry gates. These gates will be fine-tuned to ensure that they can manage the blend of non-linear character features with the linear word embeddings efficiently.

5 Conclusion

The Character-Aware Neural Language Model presents a nuanced approach to language modeling that leverages the granularity of characters to determine the context of words. The implementation could yield a model that can grasp details that might be ignored in traditional models.

References

[1] Yoon Kim, Yacine Jernite, David Sontag, and Alexander Rush. Character-aware neural language models. *Proceedings of the AAAI Conference on Artificial Intelligence*, 30(1), Mar 2016.