

COMP 550: READING ASSIGNMENT 1

JAN TIEGGES

OCTOBER 19, 2023

The Paper *Character-level Convolutional Networks for Text Classification* by Zhang et al. (2015) introduces a novel approach to text classification by utilizing character-level Convolutional Neural Networks (ConvNets). The authors demonstrate the effectiveness of character-level ConvNets by constructing large-scale datasets and comparing their performance with traditional models such as Bag of Words, n-grams, and TFIDF variants as well as deep learning models.

One strength of the paper is the empirical approach of performing an extensive comparison with multiple models and different datasets. In addition, a thorough analysis of the various factors on ConvNet performance is performed, such as the size of the dataset, the degree of curation, and the choice of alphabet. The authors themselves have held back on hypotheses regarding some of these findings and rather focused on presenting the results.

One limitation of the paper is its restriction to text classification, whereas an exploration of the generalizability to other natural language processing tasks would have been interesting. Furthermore, the influence of hyperparameters, such as momentum, is not discussed, which leaves the reader uncertain about their influence. The use of a fixed alphabet also limits the usefulness of this model for languages with more diverse character sets or texts with frequent out-of-vocabulary (OOV) items. In general, the technical details are sometimes difficult to understand without sufficient background in machine learning, as for example in the case of temporal max-pooling and its usefulness.

The authors highlight a close relationship between ConvNets and TDNNs in early deep learning research, as the latter are essentially ConvNets which model sequential data. TDNNs are intended for data with temporal structure such as speech signals and use time delay layers to capture temporal dependencies. ConvNets are more designed for grid-like data such as images and use convolutional and pooling layers for feature extraction. However, in the case of a temporal one-dimensional ConvNet, the latter can be applied to temporal sequential data, unifying the two, although they are different in their architectures.

The authors do not treat OOV elements explicitly, but instead use a fixed alphabet of characters and exclude words that contain characters outside this alphabet (quantized as all-zero vectors). This approach differs from the unknown tag and smoothing methods discussed in class, which treat OOV elements explicitly by either replacing them with an unknown-tag or adding a small probability mass to all words in the vocabulary.

In Table 4, the results of the character-level ConvNet (in different variants) are compared with many traditional and deep learning models on large datasets. For this purpose, the test error is used. Table 4 allows for a comprehensive model comparison and highlighting of the strengths and weaknesses of each method as well as dataset size and degree of curation on ConvNet performance. Although this is a valid method of presenting results, the table is a bit overwhelming at first glance due to its sheer size. It would certainly have been a valid idea to split the results into individual subgroups and, for example, separate the comparison of ConvNets with traditional models and the effect of word2vec.

$$\beta = 0.9$$

Overall, this work makes a valuable contribution to text classification by applying a completely new method using character-level ConvNets and offering insights into the strengths

and limitations of different machine learning models.