Character representations in neural contentscoring models: Benefits

Navaneeth S • 17NA30013 • navaneeth@iitkgp.ac.in

May 10, 2020

Abstract

The purpose of this paper is to introduce the reader to the benefits of character representations in neural contentscoring models and show how it helps to account for misspelling. The original paper[1] was written by Riordan, et al.

1 Introduction

By looking at the character-level compositions of words, character level embedding finds the numeric representation of them by using an one-dimensional convolutional neural network (1D-CNN). More broadly speaking, 1D-CNN is an algorithm capable of extracting information from shorter segments of a long input sequence.

Till the release of this paper, there wasn't much research on the effect of character representations in real-world scoring scenarios. We know that there is a lot of study in the neural network-based character-based representations. This paper uses this approach on educational scoring applications.

They were making use of the following techniques:

- Enriching the input representations with morphological information (Peterset al., 2017; Chen et al., 2018)
- Accounting for noise and out-of-vocabulary inputs (Luong and Manning, 2016)
- Both the above (Madnani et al., 2017)

They find two claims to account for noise in inputs which are made for employing character representations. It has been found that models which are sensitive to characters with input words having more misspelled words perform poorly. This is because the standard word-only neural models ignore these tokens. On introducing a character representation, it will be easier for the model to distinguish spelling errors which are there in the input (Horbach et al., 2017). They use the following hypothesis to operationalize this claim:

- Hypothesis 1: Models with character representation found to improve their performance when compared to models with only word representations on feeding more misspelling in the input data. This can be showed by a statistical interaction between the addition of character representations to a model and number of misspellings in the input.
- Hypothesis 2.1: However, on spell-corrected input, the performances of both models with and without character representation will be similar.
- Hypothesis 2.2: It is also true that the performance of the models with character representations will be the same when there are misspelled words in the input and when there are not. This paper test these hypotheses on content-based questions from formative and summative assignments using neural models and they are large in number.

2 Application

- This paper shows that even though the improvement in result with character representation is small but durable when used with neural models and this does not increase significantly as the number of spelling errors increases.
- They show that the improvement with spell-corrected input improves model performance more than the addition of character representations and spell correction can improve the models with additional character representations
- They achieve a new state of the art on the Automated Student Assessment Prize Short Answer Scoring (ASAP-SAS) dataset.

3 Approaches to address the problem

The network architecture they have used is given in Figure 1. Pretrained embeddings are fed to a bidirectional GRU for a word token-only model where the hidden states of the GRU includes pooling attention mechanism. Now, for a model using character representations, a sequence of 25-dimensional character embeddings are used to represent a word which is then encoded with an encoder. Before feeding, the text is preprocessed. This include spacy tokeniser and scaling in the range of 0 to 1. GloVe 100-dimensional vectors is used for word tokens. Exponential moving average method is used during training with MSE loss. Hyperparameters are tuned for both character and combined word-character embeddings. A spelling detection and correction system which employs a set of large-scale dictionaries and language models based on the approach described in Flor (2012) and Flor and Futagi (2012) was used in all experiments. Mean squared error (MSE) and quadratic weighted kappa (QWK) are used as evaluation metrics for the model. Also, generalized linear mixed-effect models (GLMMs) (Harrison et al., 2018) are used to analyze the robustness of performance improvements with character representations.

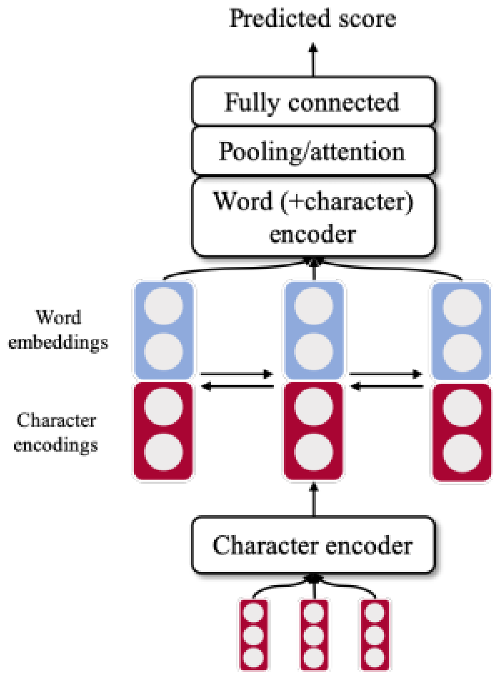

Figure 1: Neural network architecture

4 Interesting results

The model is evaluated on ASAP-SAS, Formative-K12-SAS and Summative-LAS datasets. The mean prediction error by models without spell correction on datasets are given in Figure 2, Figure 3 and Figure 4. Numbers on the bars represent the difference between word+characters and word.

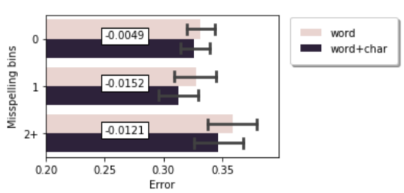

Figure 2: On ASAP-SAS dataset

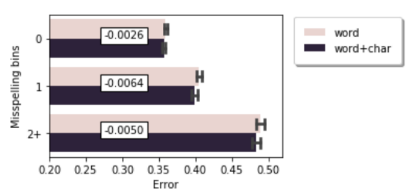

Figure 3: On Formative-K12-SAS dataset

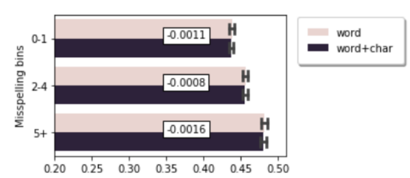

Figure 4: On Summative-LAS dataset

On ASAP-SAS dataset, we see that the models with character representations out-perform their word-only counterparts. Also, the spell-corrected models outperform the corresponding uncorrected models with the same representations. The spell-corrected model with character representations achieves the highest performance. The same trends that were observed for ASAP-SAS are observed here. Moreover, on this dataset, the mean MSE and mean QWK trends are consistent. On Summative-

LAS dataset, however, what is striking is the degree to which spelling correction improves model performance: QWK scores increase about 15 points.

5 Critical thought

This paper is the first one to discuss character representations in neural network models for automated content scoring. They have tested three hypotheses with three different datasets. This paper cannot prove that the character representations can improve the spelling variation in the input. Even though they have observed an improvement in word+character models over word-only models, this couldn't generate much statistically significant difference.

Also, they were able to show that models with character representations do not learn about the relations between spelling variation and scores. Larger training data might prove that. However, we can't find large datasets for training in educational applications.

Spelling correction will be helpful to improve the model performance of word-only models on the addition of only character representations. The data with the most spelling errors accounts for this. A new state-of-the-art result on the ASAP-SAS official test set is established by combining the spelling correction and character representations.

The results verifies that character representations on word-based neural models can provide small gains in performance consistently. Hence, character representations may provide some benefit in practice in neural models for content scoring, but that they are unlikely to serve as a replacement for spelling correction of the training data.