Riordan et al., 2019

How to account for mispellings:

Quantifying the benefit of character representations in neural content scoring models

Takehome:

"Models with character representations outperformed their word-only counterparts...lower MSE and higher QWK" p. 121

Datasets

- ASAP-SAS: 10 questions with large number of responses for each question, sentence or two in length
- Formative-SAS: dataset collected by ETS (relatively short answers)
- Summative-LAS: 20 questions, mean number of words is 230

Methods

Word only model

pretrained word embeddings into bidirectional GRU. Hidden states of GRUs are either pooled or go through an MLP attention mechanism Output of the encoder goes through sigmoid fully connected layer which produces a score

Character + word models

Each word is represented with a sequence of 25-dimensional character embeddings. "Character embeddings are concatenated with the word embeddings prior to the word-level encoder" (p. 119)

Results

ASAP-SAS

While adding character representations performed better than just spelling correction, the effect of adding character representations was not statistically significant in the GLMM model and using spelling corrections was not significant either.

No evidence for interaction between character representations and spelling correction in the GLMM.

Formative K12-SAS

Same general trend as ASAP-SAS

- character and word representations outperform word representations
- spelling corrected models outperformed non-spelling corrected models

Statistical significance between the different representations and the different methods of spelling

correction but no interaction observed between mispelling bins and the representation used.

"The difference between feature sets and between mispellings bins was significant even when controlling for score and number of words" (p. 123)

Large majority of responses had no spelling errors. 3 spelling bins used (0, 1, 2+)

Q: Is spelling not what the character representations are able to capture? Is it instead morphological variation?

- What if you ran a stemmer over the input? Would the difference between word+character embeddings and plain word embeddings go away? Surely someone has done this.

Q: I thought that the addition of character representations was helpful for two of the datasets but not the last one. The conclusion reached was that character representations were not as helpful as spelling correction but I think this was only significant for the 2nd dataset.

Q: Are the character representations alone enough? (what if you dropped words)