The Effect of Different Writing Tasks on Linguistic Style: A Case Study of the ROC Story Cloze Task: Supplementary Material

Anonymous EMNLP submission

1 Experimental setup.

This section describes the experimental setup of both Experiments 1 and 2.

In both experiments, we add a START symbol at the beginning of each sentence.[1] For computing our features, we keep n-gram (character or word) features that occur at least five times in the training set. All feature values are normalized to $[0, 1]$. For the POS features, we tag all endings with the Spacy POS tagger.[2] We use Python's sklearn logistic regression implementation (Pedregosa et al., 2011) with L_2 regularization, performing grid search on the development set to tune a single hyperparameter—the regularization parameter.

For Exp. 1, as the story cloze task doesn't have a training corpus for the *right* and *wrong* endings, we use the development set as our training set, holding out 10% for development (3,366 training endings, 374 for development). We keep the story cloze test set as is (3,742 endings).

For Exp. 2, as there are far more *original* instances than *new* instances, we randomly select five *original* sets, each with the same number of instances as we have *new* instances (3,366 training endings, 374 development endings, and 3,742 test endings). We report the average classification result.

2 Neural Language Model Training

This section describes the training details of our neural language model.

We train the LM using a single-layer LSTM of hidden dimension 512. We use the ROC stories for training,[3] setting aside 10% for validation of the language model. We replace all words occurring less than 3 times with a special out-of-vocabulary character, yielding a vocabulary size of 21,582. Only during training, we apply a dropout rate of 60% while running the LSTM over all 5 sentences of the stories. We train the LM using the Adam optimizer (Kingma and Ba, 2015) and a learning rate of $\eta = .001$, while minimizing cross-entropy.

[1] 99% of all sentences end with a period or an exclamation mark, so we do not add a STOP symbol.

[2] http://spacy.io/

[3] We use the extended, 100K stories corpus, released in winter 2017.

References

Diederik Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *Proc. of ICLR*.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. 2011. Scikit-learn: Machine learning in Python. *JMLR* 12:2825–2830.