Rule-Based Sentence Quality Modeling and Assessment using Deep LSTM Features

Pablo Rivas* **and Aishwarya Pagalla**
Department of Computer Science
Marist College
Poughkeepsie, NY 12601

1 Introduction

People trying to master the art of writing correct sentences in English for the first time often find themselves in the struggle of learning the essential rules to do so. In an effort to assist in this process we studied how to represent a sentence with the purpose of analyzing its quality according to well-known rules. We make use of recent developments in deep learning to achieve a rich representation of single sentences, commonly known as embedding [1]. Sentence quality has been traditionally evaluated in the context of machine translation to measure success [2, 3]. However, our work measures the quality of a sentence as the end product of the model. Recently, recurrent neural networks (RNNs) with long-short term memories (LSTMs) have been used for word embedding [4], sentence embedding [5], and paragraph or document embedding [6] with very good results. These works have demonstrated the utility and robustness of finding discriminative vectorial representations that preserve context and major structures of words and bodies of text. Thus, we decided to study their utility in modeling and assessing the overall quality of individual sentences.

2 Methodology and Results

2.1 Rules

There are well-known rules that make a sentence *a good sentence*. We focused on the following five:

1. **Subjects**. The subject must be the main character, not actions expressed as abstract nouns.
2. **Verbs**. The important actions in the sentence should be verbs, not abstract nouns.
3. **Introductory Phrases**. Introductory phrases in a sentence (if any) should follow rules 1 and 2 and should not be too long; around five words is acceptable.
4. **Nouns**. Strings of consecutive nouns (three or more) should be avoided to preserve sentence clarity.
5. **Conciseness**. Words that mean little or nothing, words that repeat the meaning of other words, words implied by other words, should all be avoided.

2.2 Data

These rules are nearly impossible to define in a precise manner without falling into endless exceptions due to the natural complexities that the English language produces having a context-dependent grammar. Thus, we approached the problem by creating a dataset labeled by English experts evaluating a sentence using the five rules above. The data comes from college students writing after the removal of all personal identification data. The average length of the sentences in the dataset is 110 words, and the vocabulary size is 73140.

*Pablo.Rivas@Marist.edu. Partially supported by New York State Cloud Computing and Analytics Ctr

32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada.

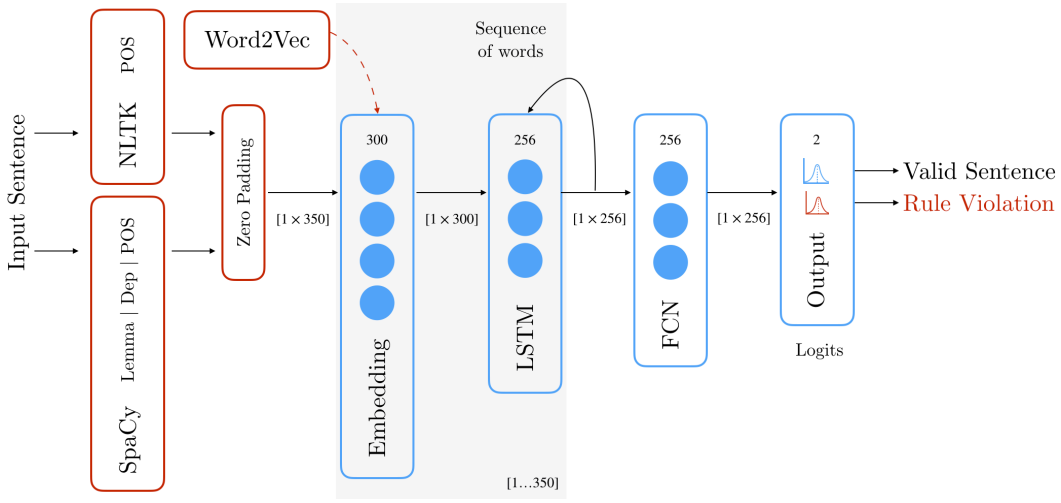

Figure 1: Deep architecture for sentence quality assessment.

Table 1: Results for each rule. Mean absolute error and accuracy are reported in standard cross validation.

Metric	Rule 1	Rule 2	Rule 3	Rule 4	Rule 5	Avg.
MAE	0.097 ± 0.01	0.109 ± 0.00	0.143 ± 0.01	0.089 ± 0.01	0.229 ± 0.01	0.133
ACC	90.45 ± 0.01	89.23 ± 0.00	85.87 ± 0.01	91.23 ± 0.01	77.43 ± 0.01	86.84

2.3 Architecture

The input to our model is a sentence of arbitrary length. Following a common practice [7], we use the SpaCy NLP library for part-of-speech (POS)-tagging, word dependency analysis, and word lemma identification. Similarly, we use the NLTK library for POS-tagging [8]. One of the reasons to use both libraries is due to their differences in the methodology for POS-tagging.

Once these features are extracted for a sentence, it follows to zero-pad before embedding. The embedding layer has 300 neurons and it is pre-trained using Word2Vec [9], which in our experiments increased accuracy by 3.2%.

The embedding layer is followed by an LSTM layer that is used to characterize and preserve any potential rule violations. This layer has 256 neural units and has a sequence length of 350. The RNN is followed by a fully connected network (FCN) with 256 neurons and an output layer with 2 neurons using logit activations. These two neurons will output a probability of the violation or fulfillment of each rule. This architecture is depicted in Figure 1. For every single rule, there is a model following the exact same architecture.

2.4 Results

The overall cross-validated average accuracy is 86.84% and the mean absolute error is 0.133 across all five rules. Table 1 shows the results for each individual rule in cross validation. The mean absolute error is calculated as $\frac{\sum_{i=1}^{N} |y_i - \hat{y}_i|}{N}$ where y_i is the i-th target output and $\hat{y}_i$ is the predicted output of the network. The rule most successfully predicted is the rule of nouns with 91.23% while the worst is the rule of concision with 77.43%. Intuitively, the rule about concision is extremely complicated to model as it needs a larger corpus relating meaning of phrases to a summarized expression of the same meaning. Further work will explore this alternative.

A live demo of this project can be found in: https://wa.reev.us in which we implemented the models to give sentence-level feedback to people learning to write good sentences.

References

[1] Hamid Palangi, Li Deng, Yelong Shen, Jianfeng Gao, Xiaodong He, Jianshu Chen, Xinying Song, and Rabab Ward. Deep sentence embedding using long short-term memory networks: Analysis and application to information retrieval. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 24(4):694–707, 2016.

[2] Ngoc-Quang Luong, Laurent Besacier, and Benjamin Lecouteux. Word confidence estimation and its integration in sentence quality estimation for machine translation. In *Knowledge and Systems Engineering*, pages 85–98. Springer, 2014.

[3] Lucia Specia, Marco Turchi, Nicola Cancedda, Marc Dymetman, and Nello Cristianini. Estimating the sentence-level quality of machine translation systems. In *13th Conference of the European Association for Machine Translation*, pages 28–37, 2009.

[4] Xuezhe Ma and Eduard Hovy. End-to-end sequence labeling via bi-directional lstm-cnns-crf. *arXiv preprint arXiv:1603.01354*, 2016.

[5] Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. Show and tell: A neural image caption generator. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3156–3164, 2015.

[6] Jiwei Li, Minh-Thang Luong, and Dan Jurafsky. A hierarchical neural autoencoder for paragraphs and documents. *arXiv preprint arXiv:1506.01057*, 2015.

[7] Tom Bocklisch, Joey Faulker, Nick Pawlowski, and Alan Nichol. Rasa: Open source language understanding and dialogue management. *arXiv preprint arXiv:1712.05181*, 2017.

[8] Steven Bird and Edward Loper. Nltk: the natural language toolkit. In *Proceedings of the ACL 2004 on Interactive poster and demonstration sessions*, page 31. Association for Computational Linguistics, 2004.

[9] Yoav Goldberg and Omer Levy. word2vec explained: deriving mikolov et al.'s negative-sampling word-embedding method. *arXiv preprint arXiv:1402.3722*, 2014.