Daniyah Arshad

CS 4395.001

23 March 2023

Controllable Dictionary Example Generation:

Generating Example Sentences for Specific Targeted Audiences

The paper being discussed in this essay is Controllable Dictionary Example Generation: Generating Example Sentences for Specific Targeted Audiences, penned by Xingwei He and Siu-Ming Yiu. Xingwei He appears to be a student studying at the University of Hong Kong in the Department of Computer Science, while Siu-Ming Yiu is a recognized professor in the same department, and seems to have been the advisor assigned to He. He has published 9 papers total, with this paper being a more recent and thus, less well known work of his. Meanwhile, Yiu has published over 300 papers, and has over a thousand citations on 4 of those papers, and other papers, though not in the thousands range, still have hundreds of citations.

The problem addressed by the paper is the issue of generating dictionary sample sentences. Often, dictionaries have example sentences to supplement an understanding of how a specific word may be used in a sentence, including multiple sentences if the word is polysemous. Generating these sentences takes many experts working together to come up with the best sentence that can best express the meaning of a word without losing clarity, and it often takes a great deal of human time and effort. Additionally, these example sentences are not always applicable to people of varying education levels; sometimes, they might be too complex for some people to understand. Using AI to automate the process is intended to solve both of these issues,

utilizing an element of human control to determine the education level for the generated sentence.

He has written 6 papers before this one, going over various concepts of lexically constrained sentence generation, optimizing and reducing sampling time for biterm sampling models, improving answer selection, and hashtag recommendation for twitter. His prior work seems to support his research over dictionary example sentence generation now, as his experience with generating lexically constrained sentences and optimizing existing processes assists in generating lexically constrained sample sentences and optimizing the time it takes to generate those sentences. Meanwhile, Yiu has countless topics and papers written before this specific paper, with enumeration of all of them being difficult. However, many of his papers go over biology, specifically RNA related topics, and computer science, specifically NLP and digital forensics related topics. His prior research over NLP likely assisted in writing and advising for this current paper.

The authors of the paper claim that while dictionary definition generation is more commonly studied, with many people having approached the subject in an effort to automate and optimize it, the concept of generating example dictionary sentences is a new concept brought forth by the authors themselves. After a cursory google search, their claim appears to be true, as there were no other results specifically about generating sample sentences for the purpose of a dictionary. As such, the entire concept of generating sample sentences for a dictionary is in and of itself the grandest unique contribution of this paper, as it highlights a new issue and resolves it.

The authors evaluated their work by answering four simple questions about their generated sentences to determine whether or not they accurately represented the purpose they were generated for. These questions are: Does the generated sample contain the targeted word?, Does the targeted word in the in the generated sentence convey the given sense?, and finally, Are the outputs fluent and diverse? With the final question counting as two separate questions, “Is the output fluent” and “Is the output diverse”, and analyzing each of these questions using human made samples as reference for answering them.

The first author, Xingwei He, has received 121 citations total, and the second author, Siu-Ming Yiu, has received 26, 945 citations total, and stands as the person with the most citations. I believe their work is important for automating a very necessary aspect of dictionaries, that being example sentences, and shortening the human labor required to write those sentences while also providing greater functionality by intending to make sentences directed at different reading levels. It would make it easier and less time consuming to construct or ask for sentences specified for a certain audience, and it would also mean that an expert doesn’t need to be available to provide an example sentence– a user looking for how a certain word can be used in a sentence can instead use AI to generate a sample sentence instead. Overall, it seems like intriguing technology with a lot of potential.