Ikhlaq Ahmad

Dr. Karen J. Mazidi

CS 4395

April 5, 2023

**Quality Controlled Paraphrase Generation: Summarized Analysis**

Paraphrasing is one of the surest ways of conveying a meaning of something in a clear, concise, and simplified way. Paraphrasing using computing machines use many techniques and algorithms to optimize the original text into a high-quality alternative. Some of these techniques include various probabilistic methods, neural networks, syntactic trees, and deep learning approaches etc. However, current paraphrasing techniques often manipulate the original context only to deliver it slightly, if not significantly, different. Therefore, it is still a challenge to produce a fully paraphrased text with almost no manipulated context and no added complexity to change its meaning even slightly.

One of the solutions has been proposed by Elron Bandel et al., who have done a thorough research under the collaboration of International Business Machine (IBM) and Computer Science Department, Bar Ilan University at Tel Aviv, Israel. They have come up with a term called Quality Controlled Paraphrase Generation, which is also the title of their research.

The authors “argue that it is difficult to preserve meaning when language variation rises, making it difficult to produce high-quality paraphrases. Newer works manage the syntactic tree and other particulars of the paraphrase to provide pleasing outcomes. Nevertheless, they lack the capacity to directly regulate the generated content's quality and have limited scalability and flexibility.” (Bandel et al. 596).

The purpose of their research is to find a way to control the quality of the paraphrased text such that it is fully optimized and controlled in terms of linguistic variation. Their suggested methodology is Quality Controlled Paraphrase Generation (QCPG) that enables for direct management of the quality dimensions in a quality-guided, controlled paraphrase generation process. Additionally, they propose a technique that, given a sentence, identifies places in the quality control space that are anticipated to produce the best possible generated paraphrases (Bandel et al. 596).

According to them: “The main component of our solution is a quality-controlled paraphrase generation model (QCPG), which is an encoder-decoder model trained on the task of controlled paraphrase generation. Given an input sentence s and a control vector c = (csem, csyn, clex), the goal of QCPG is to generate an output paraphrase QCPG (s, c) that conforms to c. We train QCPG using the training set pairs (s, t), by setting c to be q (s, t), and maximizing P(t|s,c = q(s,t)) over the training set via the autoregressive cross entropy loss.” (Bandel et al. 598). The paper argues that even if quantifying linguistic diversity is easier than quantifying semantic similarity, reliable metrics like Self-BLEU can do it automatically. As a result, the authors use automatic metrics to evaluate lexical and syntactic variance while also using human annotation to confirm the semantic evaluation.

In this study, the authors present a novel controlled paraphrase generation model that employs metrics for the generated paraphrases' quality to encourage the growth of paraphrases with the required quality. They have shown the model to have acquired a high level of control and provide a solution to the challenging problem of identifying appropriate control values. In addition to offering a simple and effective way for controlling the quality of a model's output, the quality control paradigm allows for a comprehensive view of the data, the training process, and the final model analysis.

This article was published in 2022 and has received a total of 11 citations as per the Google Scholar website. This article has been produced by Elron Bandel, Ranit Aharonov, Michal Shmueli-Scheuer, Ilya Shnayderman, Noam Slonim, and Liat Ein-Dor. Contrary to the typical lead author, Elron is not a PhD, as per his profile on the Google Scholar; however, this author has done extensive research in the field of NLP by as shown by publications which entirely focusing on the field of Natural Language Processing.

According to the [www.scholar.google.com](http://www.scholar.google.com), the number of the citation(s) each author received on all their publication(s) is listed in the table below:

|  |  |
| --- | --- |
| Author | Total Citation(s) |
| Elron Bandel | 35 |
| Ranit Aharonov | 8061 |
| Michal Shmueli-Scheuer | 669 |
| Ilya Shnayderman | 197 |
| Noam Slonim | 6920 |
| Liat Ein-Dor | 2816 |

According to the table, Ranit Aharonov has received the greatest number of citations.

In my opinion, this article establishes the foundation of generating high quality paraphrase text by utilizing a controlled approach. By employing different parameters, a desired type of paraphrasing can be achieved. This is a relatively newer approach and, based on the data shown in the results, have shown a significantly improved and accurate results.

Work Cited.

1. Bandel, Elron, et al. “Quality Controlled Paraphrase Generation.” *ACL Anthology*, IBM Research, Bar Ilan University, May 2022, [https://aclanthology.org/2022.acl long.45/](https://aclanthology.org/2022.acl%20long.45/). [Access: April 9, 2023]

Works Cited

AuthorLastName, FirstName. *Title of the Book Being Referenced*. City Name: Name of Publisher, Year. Type of Medium (e.g. Print).

LastName, First, Middle. "Article Title." *Journal Title* (Year): Pages From - To. Print.