A model that can generate humorous versions of existing headlines
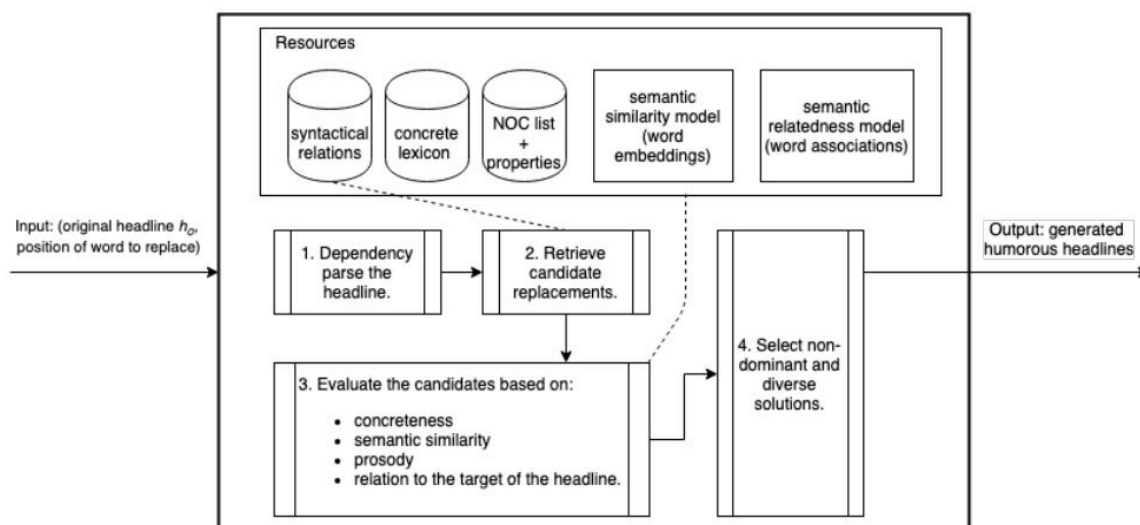
October 4 2021, by Ingrid Fadelli

Figure 1: A diagram visualizing the process of humor generation.

Diagram summarizing the process through which the model generates humorous headlines. Credit: Alnajjar & Hämäläinen.

Over the past decade or so, computer scientists have developed a growing number of computational models that can generate, edit and analyze texts. While some of these models have achieved remarkable results, some aspects of human language and communication have proved particularly difficult to replicate computationally.

One of these aspects is humor, the human ability to say or write things that are funny. Humor is a subtle and inherently human quality; thus, reproducing it in machines is far from an easy task.

Researchers at University of Helsinki have recently attempted to artificially replicate humor in machines, by developing a framework that can turn existing news headlines into humorous ones. This model, first introduced in a paper pre-published on arXiv and presented at the 12th International Conference on Computational Creativity (ICCC 2021), was trained to analyze headlines in an existing dataset and replace words in them to give them comical or amusing qualities.

"Automated news generation has become a major interest for news agencies," Khalid Alnajjar and Mika Hämäläinen, the two researchers who conducted the study, wrote in their paper. "Oftentimes, headlines for such automatically generated news articles are unimaginative, as they have been generated with ready-made templates. We present a computationally creative approach for headline generation that can generate humorous versions of existing headlines."

The recent paper by Alnajjar and Hämäläinen draws inspiration from a previous work by three researchers at University of Rochester and Microsoft Research AI, who introduced Humicroedit, a dataset containing over 15,000 annotated news headlines. In this study, the researchers identified strategies for making headlines funny that are commonly used by humans, which they found to be aligned with existing theories of humor.

The team at University of Helsinki devised a model that uses some of these strategies to change non-humorous headlines and make them more amusing for readers. To do this, it tries to find funny substitutes for some of the words in existing headlines.

Two examples of the headlines generated by the researchers' model are: "Trump eats the wrong Lee Greenwood on Twitter" and "U.S. says Turkey is helping ISIS by combing Kurds in Syria."

Humorous headline by our system	Original word	Human replacement
Thieves carry out elaborate van heist to steal millions in **cereal**, Swiss police say	cash	blouses
Trump **eats** the wrong Lee Greenwood on Twitter	tags	woos
'I was very angry' at Trump, says Myeshia Johnson, widow of fallen **sock**	soldier	cake
Trump Tried To **Climb** Heather Heyer's Mother During Funeral: 'I Have Not And Now I Will Not' Talk To Him	call	proposition
U.S. says Turkey is helping ISIS by **Combing** Kurds in Syria	bombing	feeding

Table 1: Examples of generated headlines.

Credit: Alnajjar & Hämäläinen.

To evaluate the effectiveness of their model, Alnajjar and Hämäläinen used it to change 83 headlines randomly selected from the Humicroedit dataset and make them more humorous. Subsequently, they asked reviewers on a crowd-sourcing platform to provide their feedback on whether they found the headlines generated by the model funny or not.

Overall, the researchers found that the humorous headlines produced by their model were comparable to those generated by humans on several levels. In addition, on average, they found that human evaluators sourced online considered the headlines produced by their system funny 36% of the time. If the model is improved further, it could eventually help media agencies and journalists to come up with new funny headlines for news articles.

"As the best headlines produced by our system for each original headline can, on average, reach to a human level in terms of most of the factors measured in our evaluation, an immediate future direction for our research is to develop a better ranking mechanism to reach the maximum capacity of our system," Alnajjar and Hämäläinen concluded in their paper. "Perhaps such ranking could be learned by training a long short term memory (LSTM) classifier on humor annotated corpora."

More information: When a computer cracks a joke: automated generation of humorous headlines. arXiv:2109.08702 [cs.CL]. arxiv.org/abs/2109.08702

"President vows to cut hair": dataset and analysis of creative text editng for humorous headlines. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*(2019). DOI: 10.18653/v1/N19-1012