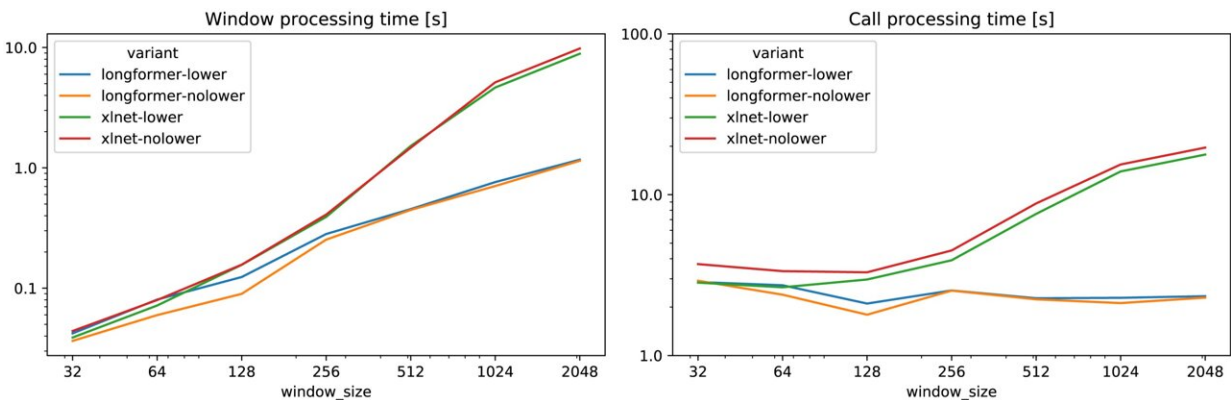


# Model moves computers closer to understanding human conversation

December 21 2021, by Wick Eisenberg



Prediction time for SWDA call sw2229 by Longformer and XLNet with different window sizes. The left-side plot shows the mean time it takes to predict a single window, and the right-side plot shows the time needed to process the full dialog. Window sizes larger than 512 imply sub-windowing for Longformer, which in this experiment has learned only 512 positional embeddings. Credit: DOI: [10.1162/tacl\\_a\\_00420](https://doi.org/10.1162/tacl_a_00420)

An engineer from the Johns Hopkins Center for Language and Speech Processing has developed a machine learning model that can distinguish functions of speech in transcripts of dialogs outputted by language understanding, or LU, systems in an approach that could eventually help computers "understand" spoken or written text in much the same way that humans do.

Developed by CLSP Assistant Research Scientist Piotr Zelasko, the new model identifies the intent behind words and organizes them into categories such as "Statement," "Question," or "Interruption," in the final transcript: a task called "dialog act recognition." By providing other models with a more organized and segmented version of text to work with, Zelasko's model could become a first step in making sense of a conversation, he said.

"This new method means that LU systems no longer have to deal with huge, unstructured chunks of text, which they struggle with when trying to classify things such as the topic, sentiment, or intent of the text. Instead, they can work with a series of expressions, which are saying very specific things, like a question or interruption. My model enables these systems to work where they might have otherwise failed," said Zelasko, whose study appeared recently in *Transactions of the Association for Computational Linguistics*.

In that paper, Zelasko adapts some recently introduced [language-understanding](#) models with the goal of organizing and categorizing words and phrases, and investigates how different variables, such as punctuation, affect those models' performance.

"We found that punctuation provides the models with very strong cues that do not seem to be otherwise present in the text, such as the content of a conversation," Zelasko said.

During his time in industry working on human-to-human conversational analytics, Zelasko noticed that many natural language processing algorithms operate well only when the text has a clear structure, such as when a person speaks in complete sentences. However, in real life, people seldom speak so formally, making it difficult for systems to ascertain exactly where a sentence starts and ends. Zelasko wanted to make sure his system could understand ordinary conversation.

"This is where the 'dialog act' framework comes in," Zelasko said. "With that, we can at least find 'units' of a conversation. This can possibly help with a large range of tasks such as summarization, intent recognition, and the detection of key phrases."

Zelasko believes that his model could eventually help companies that use speech analytics, a process that some businesses use to gain insights from analysis of interactions between customers and call center customer service representatives. Speech analytics usually involve automatic transcription of conversation and keyword searches, which Zelasko says provide limited opportunities for insight.

"With the old approach, you might be able to say that highlights of a conversation involve whatever type of phone the customer owns, 'technical issues,' and 'refund,' but what if somebody was just exploring their options and didn't actually request a refund?" Zelasko said. "That's why we need to actually understand the conversation and not simply scan it for keywords."

Zelasko predicts that his model could also someday be used by physicians, saving them valuable time they now spend taking notes while interacting with patients. Instead, a device using Zelasko's [model](#) could quickly go through the transcript of the [conversation](#), fill out forms, and write notes automatically, allowing doctors to focus on their patients.

Zelasko joined Johns Hopkins and the CLSP in January 2020, and credits the innovative and collaborative research environment for progress in his work.

"Getting these things right requires the space to let your creativity loose, and the time to digest the outcomes of your experiments, learn from them, and get it right the next time you try," Zelasko said. "But it is also important to be mindful of practical considerations and limitations when

conducting this kind of research. That's what makes CLSP a great place for this, as we have a great track record of collaboration with the industry."

**More information:** Piotr Żelasko et al, What Helps Transformers Recognize Conversational Structure? Importance of Context, Punctuation, and Labels in Dialog Act Recognition, *Transactions of the Association for Computational Linguistics* (2021). [DOI: 10.1162/tacl a 00420](https://doi.org/10.1162/tacl.a.00420)

Provided by Johns Hopkins University

Citation: Model moves computers closer to understanding human conversation (2021, December 21) retrieved 21 June 2024 from <https://techxplore.com/news/2021-12-closer-human-conversation.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.