

# Sentence Pair Modeling

## Xu creates algorithms to boost understanding, generation

With the increased information-age use of voice interfaces and chatbots, Natural Language Processing (NLP) is becoming one of the world's most significant technologies, drawing from aspects of computer science, artificial intelligence and linguistics. The goal for NLP researchers, such as Wei Xu, an assistant professor of computer science and engineering at The Ohio State University, is to develop deep-learning techniques to enable computers to process or "understand" natural language in order to perform tasks like language translation and question answering.

For a computer, fully understanding and representing the meaning of spoken or digital language is extremely difficult. Human language is very complex, often carries encoded messaging and varies based on locale, culture and generation.

"We focus on designing algorithms for learning semantics from large datasets for natural language understanding and generation, in particular with stylistic variations," Xu said.

"My research lies at the intersections of machine learning, natural language processing and social media."

One of Xu's most recent research project, leveraging Ohio Supercomputer Center (OSC)

services, involves developing a complex algorithm that evaluates sentence pair modeling, which is a fundamental technique underlying many NLP tasks. This measures the degree of equivalence in the underlying semantics of paired snippets of text; identifying whether two sentences express the same meaning; recognizing whether a hypothesis can be inferred from a premise; determining how well candidate sentences answer questions; and matching passages with an appropriate question and signaling the text region that contains the answer.

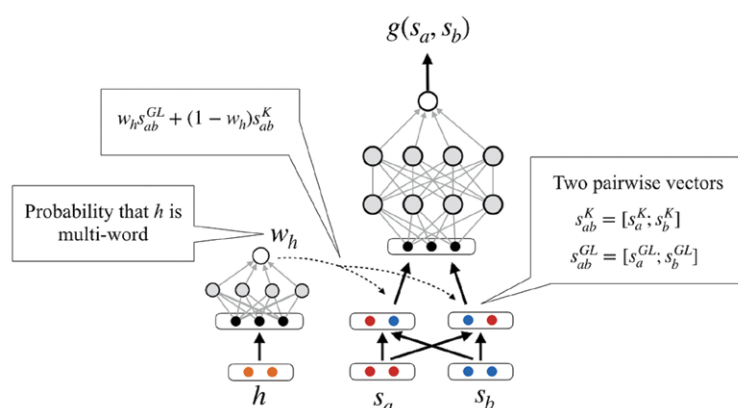
"To answer these questions and better understand different network designs, we systematically implemented five representative neural models and their variations on the popular PyTorch deep learning platform to analyze for sentence pair modeling," Xu explained.

In addition to using a small in-house GPU server to prototype their algorithms, her research team leveraged CPU and GPU nodes on OSC's Owens Cluster to process much larger modeling runs.

"We focused on identifying important network designs and presented a series of findings with quantitative measurements and in-depth analyses, releasing our implementations as an open-source toolkit to the research community."

Other research in Xu's lab involves NLP algorithms that match a complex sentence with a simpler version with nearly identical meaning, and, most recently, one that automatically divines the most likely meaning of a social-media hashtag; for example, quickly interpreting the hashtag #wingsofstrength as "Wings of Strength." •

**PROJECT LEAD** // Wei Xu, Ph.D., The Ohio State University  
**RESEARCH TITLE** // Neural network models for paraphrase identification, semantic textual similarity, natural language inference, and question answering  
**FUNDING SOURCES** // National Science Foundation, Defense Advanced Research Projects Agency  
**WEBSITE** // [cocoxu.github.io](https://cocoxu.github.io)



**FIGURE** // Wei Xu and her colleagues are working to help computers better "understand" natural language through the development of algorithms for learning semantics. This chart outlines how multi-task learning is accomplished.