

Machines just revealed the evolution of language

August 7 2017

Machine learning scientists at Disney Research have developed a new innovative model that uncovers how the meanings of words change over time.

Dr. Robert Bamler and Dr. Stephan Mandt developed the dynamic word embeddings [model](#) by integrating [neural networks](#) and statistics used in rocket control systems. The result is a complex machine learning algorithm that automatically detects semantic change throughout history.

"Promoting innovation in machine learning is a vital part of Disney Research," said Markus Gross, vice president of Disney Research.

The authors will present the model on August 9 at the International Conference on Machine Learning (ICML) in Sydney.

Their dynamic word embeddings represent semantic change over time through so-called semantic vector spaces. Words of similar meanings appear close to one another and reveal each other's relationships. Changes in these meanings appear as movements through the semantic space.

"Dynamic word embeddings could revolutionize how we investigate the evolution of language use," Mandt said. "The model uncovers otherwise latent semantic changes that become palpable through new dynamic visualizations."

For instance, the model automatically discovered that the meaning of the word 'peer' drastically changed in the last 200 years. "In the past, a peer was an earl or viscount," Bamler said. "Today, it's someone of equal standing. That is practically the opposite." Other words whose meaning changed dramatically included 'computer,' 'radio,' and 'broadcast'.

Bamler and Mandt integrated neural networks with classical statistical models. While having neural networks work on their data sets, the researchers used a Kalman filter for their transition model. It is a statistical model originally developed to trace, for instance, the position of rockets.

"Previous approaches produce different word vectors every time they are trained," Mandt said. "Every output used to be intrinsically correct, but these outputs could not be compared to detect changes over time."

Added Bamler, "We developed a mechanism to train the model on a vast corpus without splitting the data into time slices. That way, we were able to leverage a very large data set while still allowing the word vectors to drift smoothly over time."

The combination of classical statistical models with neural networks is a progressive approach able to prompt applications beyond the scope of word embeddings. To test their model, the researchers drew on a variety of data sets including 230 presidential speeches and millions of digitized books from over 200 years, as well as several months of public social media posts. In all experiments, the new model was consistently better in predicting missing words from these sources of text than existing models, confirming the validity of the observed changes.

Beyond tracing language history, Bamler and Mandt also analyzed contemporary changes in language usage. Training their model on [social media posts](#), they detected changes of word associations that reflected

current events.

Combining creativity and innovation, this research continues Disney's rich legacy of leveraging technology to enhance the tools and systems of tomorrow.

Provided by Disney Research

Citation: Machines just revealed the evolution of language (2017, August 7) retrieved 25 September 2023 from <https://techxplore.com/news/2017-08-machines-revealed-evolution-language.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.