

New technique to control offensive suggestions on web

Scientists at Microsoft and IIIT, Hyderabad, have proposed a mechanism to make browsing safer, especially for children

R. RAVIKANTH REDDY
HYDERABAD

Can offensive, sexual, racist and violence related suggestions thrown up during a search on the web be controlled?

A joint research by scientists at the Microsoft and the International Institute of Information Technology (IIIT), Hyderabad, promises to control inappropriate suggestions by search engines making web browsing safe for everyone, particularly children and women.

The team of Harish Yenala, research student in IIIT Hyderabad; Manoj Chinakotla, Senior Applied Scientist, Artificial Intelligence and Research, Microsoft India; and Jay Goyal, Principal



Safe search: Deep Learning involves building artificial neural networks trained to mimic the behaviour of the human brain.

Development Manager, Microsoft, India, have proposed a promising technique for automatically identifying such suggestions based on a new field of computer science research

known as Deep Learning (DL). This aims to build machines that can process data and learn in the same way as our human brain does.

Dr . Manoj says the DL essentially involves building

artificial neural networks trained to mimic the behaviour of the human brain. These networks can learn to represent and reason over the various inputs given to them.

The research was presented at the recent Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD) 2017 held in Seoul where it received the best paper award from among the 450 papers presented.

The DL architecture that the team proposed is called "Convolutional Bi-Directional LSTM (C-BiLSTM)", and is a combination of the strengths of both the Convolution Neural Networks (CNN) and the Bi-directional LSTMs (BLSTM).

Given a query, the C-BiLSTM uses a convolutional layer for extracting feature representations for each query word, which is then fed as input to the BLSTM layer.

This captures the sequential patterns in the query, and outputs a richer representation encoding them. This representation then passes through a fully connected network that predicts the target class before giving the output suggestion.

Reasoning ability

For example, Dr. Manoj says a child trying to search "kite" can see suggestions like "killing," using the first two letters. Socially offensive suggestions can actually lead to

confrontation. "Our research outcome is sensitive and gives only appropriate suggestions."

The advantages of the C-BiLSTM include the fact that it doesn't rely on hand-crafted features, is trained end to end as a single model, and effectively captures both local and global semantics.

The technology is being used in Microsoft's search engine Bing and will be launched soon as Application Programming Interface in Microsoft Cognitive Services.

The team is sure that the new architecture will be highly effective in online platforms such as chatbots and autonomous virtual assistants.