Keynote Lecture-1

Neural Models of Text Normalization for Speech Applications

Richard Sproat
Google, New York
Joint work with Navdeep Jaitly, Kyle Gorman and Ke Wu

In this talk, I will present our recent research on applying attention-based RNN's to the problem of text normalization for speech applications, in particular text-to-speech. In this task, we want to transform an input text, e.g. "A baby giraffe is 6ft tall and weighs 150lb", into a sequence of words that represents how that text would be read, e.g. "a baby giraffe is six feet tall and weighs one hundred fifty pounds". The state of the art for the complete text normalization problem to date still uses the 20-year-old technology of rule-based weighted finite-state transducers.

We show that RNNs provide an attractive data-driven approach, but that there are issues with it producing the occasional unacceptable error, such as reading "£" as "euro". We propose some possible solutions to those errors, and discuss future directions for this research.

D S Sharma, R Sangal and A K Singh. Proc. of the 13th Intl. Conference on Natural Language Processing, page 177,
Varanasi, India. December 2016.