2018-00288 - [CORDIS-MULTISPEECH] - PhDThesis: Natural language processing: adding new words to a speech recognition system using Deep Neural Networks

Context

Title:

Natural Language Processing: adding new words to a speech recognition system using Deep Neural Networks

- Location: INRIA/LORIA Nancy Grand Est research center France
- Research theme: Perception, Cognition, Interaction
- Project-team: Multispeech
- Scientific Context:

Voice is seen as the next big field for computer interaction. The research company Gartner reckons that by 2018, 30% of all interactions with devices will be voice-based: people can speak up to four times faster than they can type, and the technology behind voice interaction is improving all the time.

As of October 2017, Amazon Echo is present in about 4% of American households. Voice assistants are proliferating in smartphones too: Apple’s Siri handles over 2 billion commands a week, and 20% of Google searches on Android-powered handsets in America are done by voice input.

The proper nouns (PNs) play a particular role: they are often important to understand a message and can vary enormously. For example, a voice assistant should know the names of all your friends; a search engine should know the names of all famous people and places, names of museums, etc.

An automatic speech recognition system uses a lexicon containing the most frequent words of the language and only the words of the lexicon can be recognized by the system. It is impossible to add all possible proper names because there are millions proper names and new ones appear every day. A competitive solution is to dynamically add new PNs into the ASR system. The idea is to add only relevant proper names: for instance if we want to transcribe a video document about football results, we should add the names of famous football players and not politicians.

In this study, we will focus on the problem of proper names in automatic recognition systems. The problem is to find relevant proper names for the audio document we want to transcribe. To select the relevant proper names, we propose to use an artificial neural network.

Assignment

- Missions:

We assume that in an audio document to transcribe we have missing proper names, i.e. proper names that are pronounced in the audio document but that are not in the lexicon of the automatic speech recognition system; these proper names cannot...
The goal of this PhD Thesis is to find a list of relevant OOV PNs that correspond to an audio document and to integrate them in the speech recognition system. We will use a Deep neural network to find relevant OOV PNs. The input of the DNN will be the approximate transcription of the audio document and the output will be the list of relevant OOV PNs with their probabilities. The retrieved proper names will be added to the lexicon and a new recognition of the audio document will be performed.

Main activities

- Missions
  During the thesis, the student will investigate methodologies based on deep neural networks [Deng2013]. The candidate will study different structures of DNN and different representation of documents [Mikolov2013]. The student will validate the proposed approaches using the automatic transcription system of radio broadcast developed in our team.

- Bibliography:

Skills

- Skills and profile: Master in computer science, background in statistics, natural language processing, experience with deep learning tools (keras, kaldi, etc.) and computer program skills (Perl, Python).

Benefits package

- Subsidised catering service
- Partially-reimbursed public transport

Remuneration


Monthly salary after taxes: around 1596,05€ for 1st and 2nd year. 1678,99€ for 3rd year. (medical insurance included).