

Application of Neural Network Based on Speech Recognition

Functional Design Document

Group name

ElPsyCongroo

Team member:

张跃飞 11713021

向昕昊 11712617

Data selection,

Some open source datasets:

- ✓ THCHS30 Tsinghua mandarin corpus
- ✓ Free ST Chinese Mandarin Corpus

Model selection,

To get pinyin from voice, we need one model of following models:

- ✓ GRU + CTC
- ✓ CNN + LSTM
- ✓ RNN-T
- ✓ LAS
- ✓ DFSMN(Alibaba)
- ✓ Attention + LSTM + trigram language model(Xiao mi)

Performance goal

We hope to achieve at least 60% CER (Character Error Rate)

Provided functions

We want users to have better experiences on their computer that he does not need to type some duplicate content repeatedly.

At first, user should define some pairs, which has name and content. And if user say something match the name, our application will automatically input the corresponding content into the cursor's location.

For example, he does not need to type the web site into the browser. In addition, he also does not need to type the e-mail address every time he is filling in the form, where he may usually forget the specific address and has to look up it in his notebook, which is a waste of time.

We know that in some Web Brower like Chrome, it already has such a function that it stores the specific data like username, password, zip code so that user do not need to type it again next time. However, we still regard our application as competitive mainly for TWO reasons. Firstly, our users have more flexibility to use our application. User could define

their own pairs to extend its ability. Secondly, since our application is could be use in any other applications which need keyboard input, it can be executed once user need to input some content in the database. For example, they can use it when they are using Microsoft Word, PowerPoint or Excel, or in the login interface of a game.

Here is some use example:

When the user says: “qq 邮箱”, It will input 11111111@qq.com in the input text where the cursor is.

When the user says: “南科大官网”, It will input www.sustech.edu.cn in the input text where the cursor is.

When the user says: “实验室服务器”, It will input ssh [user@10.20.30.300](ssh:user@10.20.30.300) in the input text where the cursor is.

Such a match pair is totally defined by user. He is supposed to store the matching in his own computer. Then he can use it each time he executes the application.