Commonly Used Text Input Based on Speech Recognition

11712617 向昕昊

11713021 张跃飞

Overview

- Model
- Experiment
- Application

Model

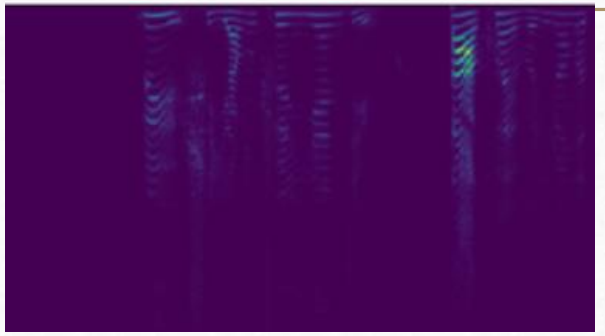
Speech Model

- VGG(CNN) + CTC
- From .wav file to Pinyin

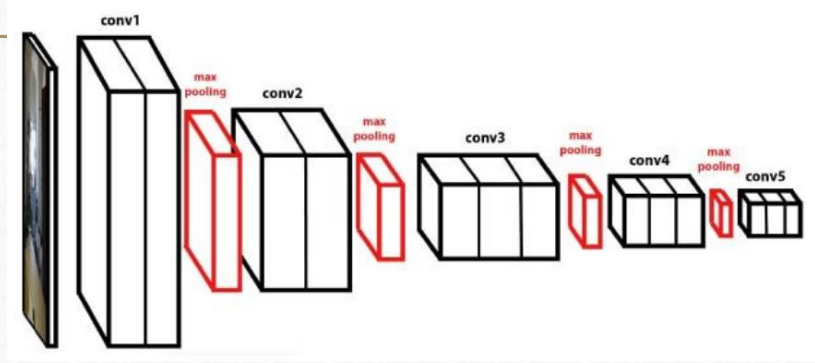
Language Model (Pinyin -> Chinese character**)**

- Hidden Markov Model with Maximum Entropy Based on Probability Graph

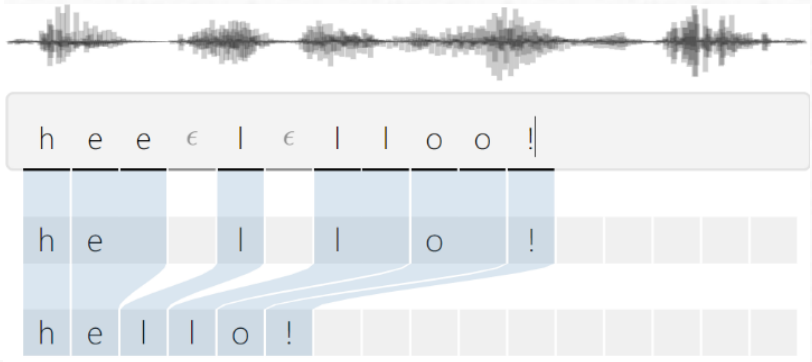
Speech model



feature extraction



Training acoustic model



CTC decoding

.wav file → Pinyin

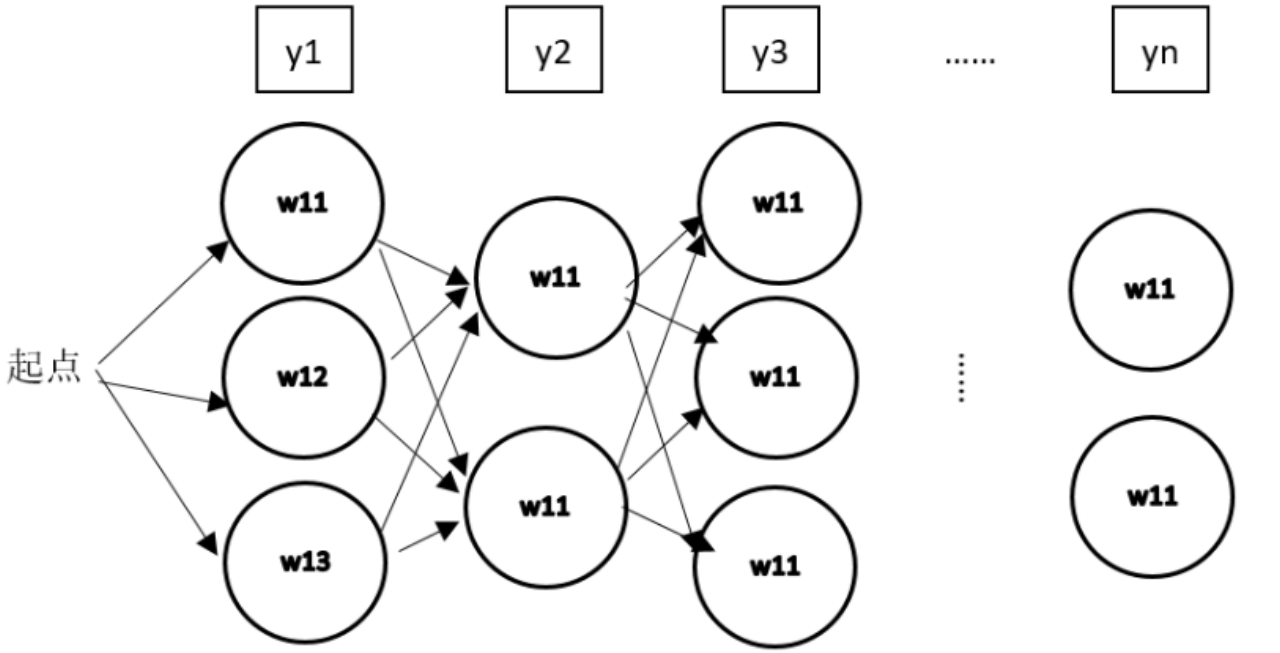
Language model

$$P(S) = P(w1,w2,\ldots,wn) = P(w1)*P(w2|w1)*P(w3|w1,w2)\ldots P(wn|w1,w2,\ldots,wn-1)$$

$$P(S) = P(w1,w2,\ldots,wn) = P(w1)*P(w2|w1)*P(w3|\ w2)\ldots P(wn|wn-1)$$

$$P(wi|wi-1) = P(wi-1,wi) / P(wi-1)$$

$$P(wi|wi-1) = \#(wi-1,wi) / \#(wi-1)$$

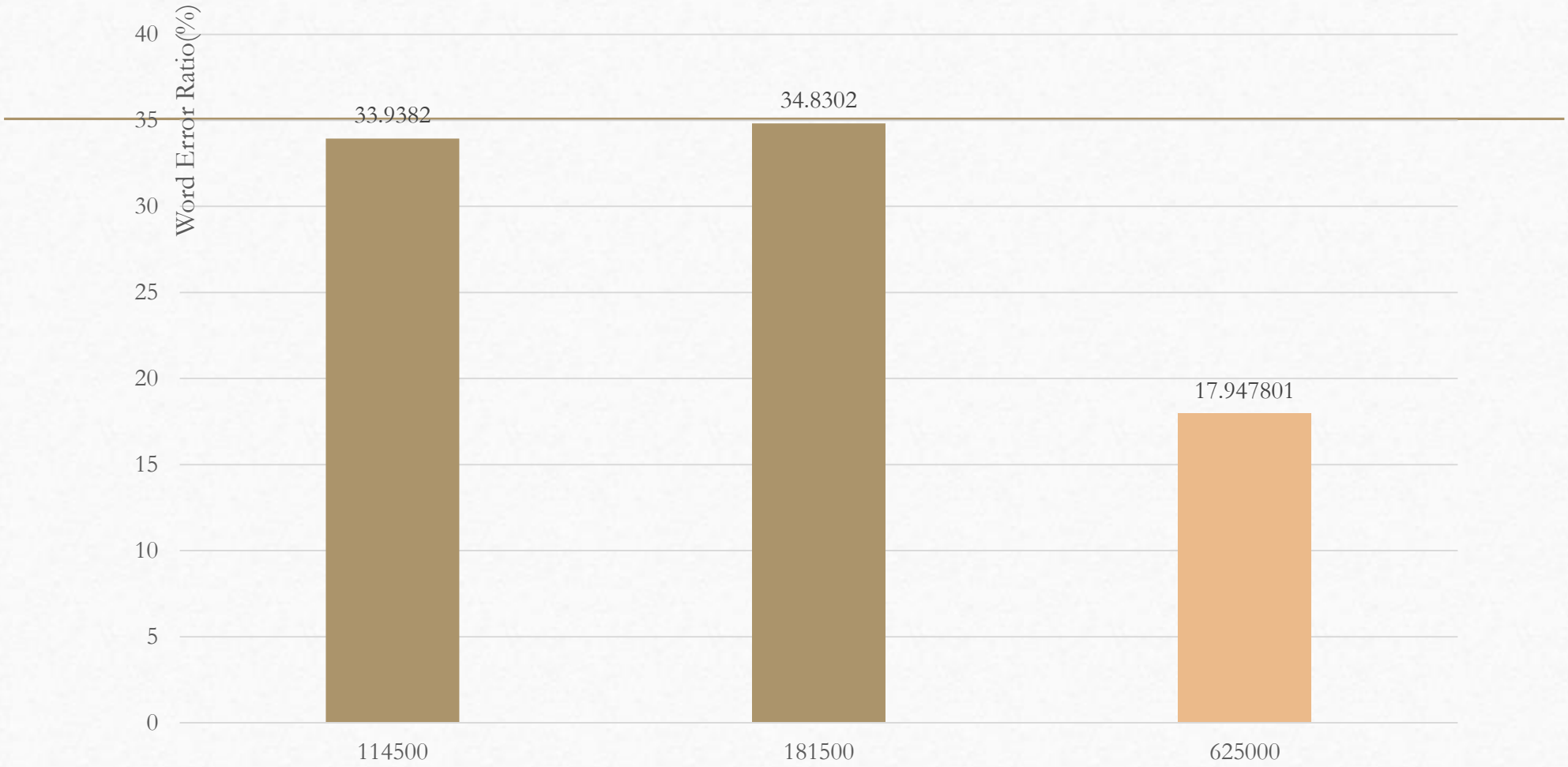


Pinyin → Chinese characters

Experiment Setting

- Software
  - Keras + Tensorflow
- Hardware
  - GTX 1080 Ti 8G
- Dataset
  - THCHS30 Tsinghua mandarin corpus
  - Free ST Chinese Mandarin Corpus

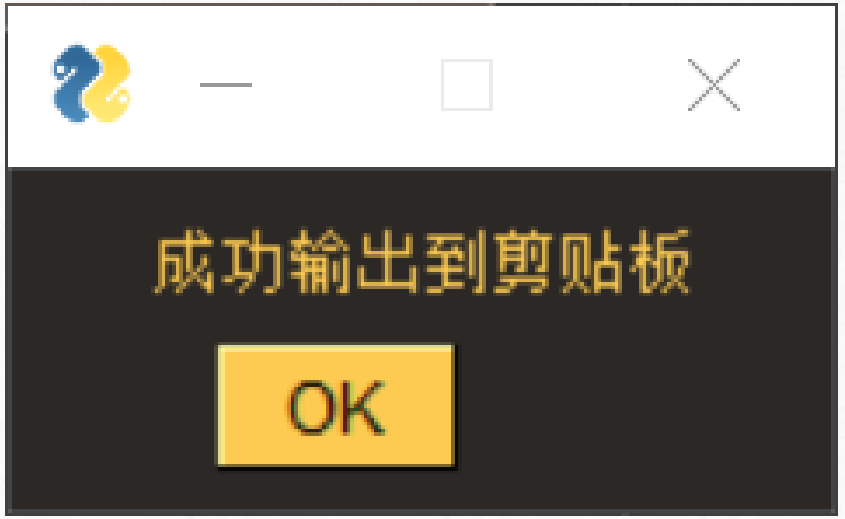
Test Results



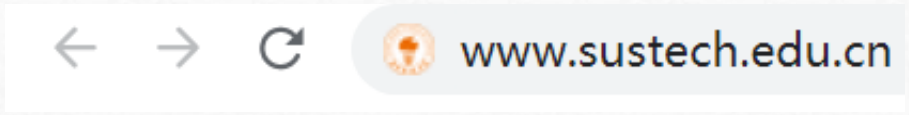
Our application

1. Input tags and triggered events

2.Do the sound recording



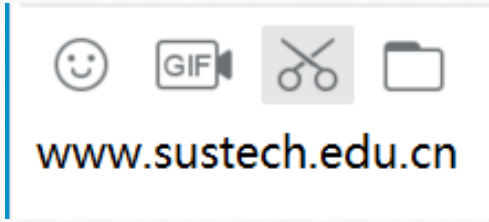
3.Use it in any text box



Country / region: **China**

Name of your academic institution:

Website of your academic institution: www.sustech.edu.cn




```
(base) PS C:\Users\ThinkPad> ping www.sustech.edu.cn
```

4.Delete when not use it

Features of our application

- User friendly
  - User Defined Matching
  - Application Scalability
  - Local Database

Use speech recognition to find output a MATCHING between user-defined label and text output

> How to use it……. Depends on your imagination!

Reference

- https://github.com/nl8590687/ASRT_SpeechRecognition

Thank you

11712617 向昕昊
11713021 张跃飞