Individual Report on Project (32): Voice Controlled Calculator
Name: Yueqi Wang
email address: yueqiwan@usc.edu
Rest of my group: Wenzheng Li, Jiyang Zhou
Name of the Bitbucket repository for your group's project: VoiceCalculator

1. Project Overview

The goal of our project is to build a voice control calculator.First, the calculator will receive a audio input said by a specific speaker, and use the speech recognizer to recognize the word of the input, and translate the waveform input to a utterance text input. Second, the calculator uses a natural language system to transfer the utterance file to mathematical expression and calculate the result. Last, the calculator uses a speech synthesis system to say the result and expression out through the calculator. We use the mobile as our calculator platform, and I am in charge of the speech recognition part, the speech recognizer will receive the voice input from the calculator. However, because the recognition system is a word-isolated and speaker-dependent recognition system, speaker need to say one word each time and it can only recognize my voice. After my speech recognition system recognize the input from mobile, the natural language system will transfer it to a expression. For example, my recognizer recognize the input is “three plus one multiply four”, then it will be transferred as “3+1*4” and calculate the result as 7, And my partner Jiyang is responsible for this part. After the expression is calculated, a speech synthesis system will let the calculator say the expression and result out. My partner Wenzheng is in charge of this part. He used Festival speech synthesis system to complete this part. Based on our evaluation, the speech recognition system’s accuracy is almost 100% due to the input corpus is relatively basic. However, there is several drawbacks during the test. First,the speaker must be myself, which is

the part need improvement in the future. In addition, the time during the input to the output is relatively long, so we need to make some change to improve the efficiency of this calculator.

2. My Project Tasks

My task is to create a word-isolated, speaker-dependent speech recognizer tool. This tools could recognize number word zero to ten and operator word: "plus",”minus","multiply","divide" said by a specific speaker. All the file for the speech recognition part is in the “recognizer” folder. The steps below is the basic steps to develop a recognizer.

1).Data Preparetion

I record the training data and test data all by myself. First, I use the HSLab tool in the HTK tool kit to record the some training file and forward to the training and recognition process. However, because I can only use HSLab at my virtual box Linux Ubuntu environment, the audio file’s quality is somehow effected by the system and it couldn’t recognize some test file that I record under IOS platform. So I decide use another recoding tool, which is Audacity, under IOS platform. I used Audacity for to make the .wav file, which is the audio file, and .lab file, which is the label file. Also, because the format of label file made by Audacity is different with the label file HTK requires, so I write a python script to transfer it to the right format. The python file is store in recognizer/data folder named “labProcess.py”. All the audio file and label file are stored in “recognizer/data” folder.

2).Acoustic Analysis

Because HTK tools cannot deal with audio file directly, so I need to first transfer all the audio file into the corresponding decoded file, the format for the decode file is .mfcc file using HCopy tool in HTK toolkit. The transferred .mfcc file are store in “recognizer/data” folder.

3).Initialization

To begin the training process, first we need to make a prototype file for each word, the prototype file contains:

number of states

form of the observation functions (associated with each state)

disposition of transitions between states

Then we can initialize using .mfcc files, .lab files and prototype file by HInit tool in HTK toolkit and generate the initialized hmm in model/hmm0 folder.

We also need to use HCompv tool to create a variance floor macro file which contains global variance vector multiplied by a factor, this file is called “VFloors”, which is in model/hmm0Flat folder.

4).Training Process

After all the preparation, we can begin out training process. I used the the model file in hmm0 folder and vFloors file that just been created, and the label file and .mfcc file, train the model iteratively for several times. At last, we can get the fully trained model.

5).Dictionary and WordNet

After the training process, we need a dictionary and a word net to let the recognizer know which word need to be recognize and the relation of the word. We first create a grammar file that contains the logical relation for each word. Because this is a word-isolated speech recognizer, so the grammar is simple as the word with preceding and trailing silence. Then I create the word dictionary and word net using HParse tool in HTK toolkit.

6).Recognition Process

First, we need to transfer the audio test file to the .mfcc file using acoustic analysis, then we can use the dictionary file and model file to recognize the .mfcc file using HVite tool in HTK toolkit. After

recognition, HVite will generate a .mlf file named recog.mlf containing the recognized word in the file. recog.mlf is stored in “recognizer” folder.

7).Evalueation

I used the HResults tool in HTK toolkit to evaluate the recognition result. Basically, the tool compares the label file made by myself with the .mlf file that contains the word recognized by the recognizer. I tested 4 audio file for each word, and the recognition accuracy is 100%. I think the accuracy is high because the word corpus is relatively simple, and the speaker is just myself, also I test the recognizer in a very quiet place. So the accuracy might drop when I use it in a quite noisy place.

8).Furture Work

There are several improvement that need to be done in the future. First, I will add more word to expand the word corpus. So the speaker is not necessarily only say number words and operator word, they can say other words as well and the recognizer could recognize it. Second, I will change it to a speaker-independent recognizer that every one could use it. Third, I will change it to a continuous speech recognizer so that the speaker doesn’t need to say one word each time. They can say a whole sentence.

3. Project Resources

- For speech recognition part,we use Audacity(http://audacityteam.org/) to record and label the data and we use HTK Toolkit(http://htk.eng.cam.ac.uk/) to train and recognize data
- For speech synthesis part, we used the Festival speech synthesis system(http://www.cstr.ed.ac.uk/projects/festival/),American

female voice installed from CMU Flite(http://festvox.org/flite/),Python library
- num2words(https://pypi.python.org/pypi/num2words)

4. References

- Nicolas Moreau(02.02.2002)-HTK (v.3.1): Basic Tutorial(Document available here http://agp1.hx0.ru/arts/HTK_basic_tutorial.pdf)
- htkbook(available under folder voicecalculator/recognizer/recourse folder)

5. Review 1

The group number was: **9**

What was the goal of the group's project?
Predict the tags(a,k,a,keywords, topics, summaries), given only the question text and its title.

How did the group attempt to accomplish the goal?
First, they did the data cleaning by processing punctuation removal, HTML tag removal, stop word removal, lexicon creation and duplicate removal.
Second, they did the Feature Extraction, they used bag of words representation, TF-IDF and part-of-speech tags method.
Third, they did the model selection by trying different models such as Baseline model, CRF with different window size, SVM with different window size, Adaboost, Tag-keyword co-occurrence model, fuzzy nearest neighbor search model and ACT-R inspired Bayesian model to select the best one.

How did the group evaluate their work and what was the result?

They used precision, recall and F1-score for the evaluation.

6. Review 2

The group number was:**11**

What was the goal of the group's project?
Disease diagnosis based on the interaction between user and the system through a chat-based approach

How did the group attempt to accomplish the goal?
First, they use a dialog management system which use slot filling/form filling thechnique to interact with users.
Second, they use a information extraction system which involves stop word elimination, spell check, sematic and sentiment analysis, stemming and lemmatization and synonym extraction.

How did the group evaluate their work and what was the result?
They evaluate their word by the performance is based off the difference form the baseline system. Two evaluation metrics are used: a subjective one based on user's feedback and a objective one based on performance statistics.

7. Review 3

The group number was:**17**
Using sentiment analysis to predict stock market

What was the goal of the group's project?
Predict the stock information by the information from social network

How did the group attempt to accomplish the goal?
They get the data form some social network like Twitter APT and YAHOO stock data,
Then they use a sentiment analysis to determine the positive or negative attitude and then compare this to the stock data.

How did the group evaluate their work and what was the result?
They evaluate their result by the accuracy. With SVM and linear regression model, their accuracy is 43.90%.

8. Review 4

The group number was:**16**

What was the goal of the group's project?
determine the category for the news dialogs

How did the group attempt to accomplish the goal?
They used text classification and parsed for the original HTML file and get the topic,
Then they use TF-IDF calculate weight for each word
They used NB, scikit-learn and SVM for the data training

How did the group evaluate their work and what was the result?
They calculate Precision, Recall and F-score as their evaluation rule.

9. Review 5

The group number was:**13**

What was the goal of the group's project?
Build a toolkit for language detection

built a binary classifier that can detect between English and Frech

How did the group attempt to accomplish the goal?
Using NLTK, they built a model to handle muti-language detection.(English, Frech, German, Swedish)
Naive Bayes binary classifier to detect between English and Frech sentences/documents using words as features
for multiple one, they used n-grams and distance measure.

How did the group evaluate their work and what was the result?
They used precision, recall and F-score, for English, the precision is 98.52%, recall is 98.93%, and the F-score is 98.72%.

10. Review 6

The group number was:**22**

What was the goal of the group's project?
Predicating a review's rating, predict exact rating(1-5 star)

How did the group attempt to accomplish the goal?
They used SVM and MegaM toolkit as the classifier.
they used a basic set of features and tried other advanced feature sets to see the improvements.

How did the group evaluate their work and what was the result?
They used accuracy to evaluate their work.

11. Review 7

The group number was:**40**

What was the goal of the group's project?
Classify cuisine based on ingredients.

How did the group attempt to accomplish the goal?
They used 37994 recipes and 6715 ingredients and cleaned the data by removing accents and special characters and used TF-IDF Frequency as the features to classify.

How did the group evaluate their work and what was the result?
They used accuracy to evaluate their result. The results show the preprocessing step reduces the execution time.

12. Review 8

The group number was:**24**

What was the goal of the group's project?
Use error detection and correction to deal with homophone confusion.

How did the group attempt to accomplish the goal?
First they created feature sets: A feature for two tokens previous to the label token, feature for POS Tags of the two tokens previous to the label token and a feature marking whether the label token is the first word in the sentence. Then they use SVM and perceptron to classify the data.

How did the group evaluate their work and what was the result?
They used accuracy and F1-score to evaluate their result. The Accuracy for SVM is 91.23%, for perception is 94.61%. F1-score for SVM is 0.8332 and for perceptron is 0.8587.

13. Review 9

The group number was:**36**

What was the goal of the group's project?
predict rating(1-5 stars) yelp reviews

How did the group attempt to accomplish the goal?
They used skewed Classes and split the reviews base for the preprocessing
Then they train the models for Doc2VexModel and n-Grams Model
Feature Types: Unigram TF
Algorithms: Naive Bayes Logistic Regression, SVM, Random Forests

How did the group evaluate their work and what was the result?
use Mean Square Error instead of Precision/Recall/F-score, The result is MSE=0.772. Average F-Score is 53.921%.

14. Review 10

The group number was:**43**

What was the goal of the group's project?
to predict a Restaurant rating based on user-generated reviews from Yelp

How did the group attempt to accomplish the goal?
They preprocessed the data by removing punctuation and extra spaces, converting words to lower case, removing stop words and

stemming. They used average perceptron, support vector regression and MegaM to classify the data.

How did the group evaluate their work and what was the result?
Extracted 25000 restaurant reviews with equal count for each rating
Trained the model with 75% training data and remaining 25% as development data for Average Perceptron
They used 5-fold cross-validation method to evaluate the model for both SVR and MegaM toolkit