Spoken Language Understanding and Intent Classification

Haoran Wang, Yi Yang, Chongxin Luo
Graduate Student
Department of Electrical and Computer Engineering
NetId: hwang293, yiyang4, cluo5

1 Introduction

Spoken Language Understanding(SLU) system has been studied and developed extensively in the recent years. Many leading technology companies have developed their own SLU system based products, such as Siri, Cortana, and Google Assistant. A SLU system consists of domain identification, intent classification and slot filling three components, and they are usually being modeled separately. [3] Most current SLU systems can identify domain and intent of individual spoken sentences very well, however, it lacks of the ability to relate sentences with previous conversations.

2 Objective

Understanding a sentence based on previous conversations is a very important task in communication. In this project, we are going to explore part of the SLU system, specifically, we are targeting to achieve sentence intent classification using machine learning. And in addition, we are going to experiment different engineering structures which helps the algorithm to understand a sentence based on previous conversations.

3 Related work and proposed approach

There are many literature works that focused on the area of SLU, the specific tasks of domain identification and intent classification are being considered as basic classification problems. Linear classifications, such as Support Vector Machines are being widely used on the task of domain identification and intent classification. [1] Moreover, Neural Networks have also being applied to this tasks, which gives a more complex model and better performance comparing with single linear classification. [4] In this project, we propose to develop a spoken sentence intent classification algorithms using Neural Network approach. The algorithm will be trained and tested on on-line text intent classification databases, and being demonstrated using Google speech recognition API. In addition, we are going to experiment with the task of classifying intent of a spoken sentences related to previous conversations. A model with Recurrent Neural Networks(RNN) will be used for this specific part of the project, [2] and by doing so to achieve a relatively positive result on understanding spoken language and intent classification.

4 Recent Progress

For the sentence intent classification problem, we used the ATIS dataset to train and test our algorithms. The ATIS dataset has been widely used in SLU system researches. It contains thousands of sentences that are classified into several categories. The goal of this project is to experiment several algorithm implementations on the sentence intent classification. Currently, we have implemented the SVM multi-class classification with two different sentence feature extraction methods, and RNN with linear regression classification.

For the SVM multi-class classification implementation, we experimented two sentence-feature extraction method. The first method, for each given sentence, we extract the first fifteen words from the sentence, and transfer each word into a integer according to the ASCII number for each character in the word. The ASCII number for each character is padded onto each other, therefore, the order of letters in the word is preserved. If a sentence has less than fifteen words, zeros are being padded onto the feature. With this feature-extraction implementation, each sentence is being transferred into a 1X15 array of integers, and it is being used as variable array for SVM algorithm to train and evaluated on. The SVM algorithm from sklearn package was being used. A total of 4000 sentences were used to train the SVM model, and

500 sentences were used to evaluate the model. There are total of 15 classes in the testing dataset, and the algorithm is able to make predictions with 5 classes. The overall accuracy for this model is 76.4%. However, with further inspection into the test result, we see that the classifier classifies majority samples into a single class, might due to the majority data in the data set is from this same class. The learning rate and overall performance of the model is shown in the figure below.

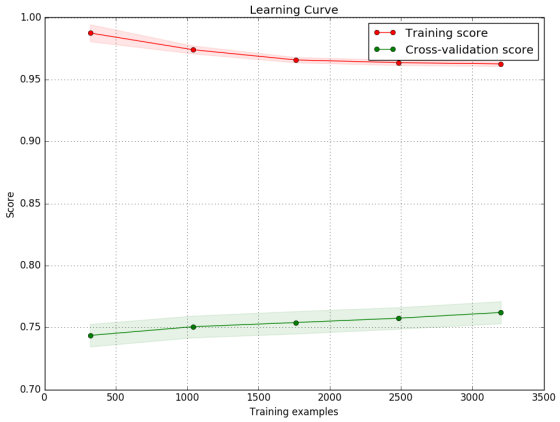

Figure 1: SVM Learning rate with first feature extraction

For the second sentence feature extraction, we used a bag-of-words combined with n-gram implementation to extract features from each sentence. Initially, we iterate through the entire training set, and stored every vocabulary in the training set into a dictionary with a paired ID number. Then a vector with length equals to the number of vocabularies is constructed for each sentence. In this vector, if a vocabulary appeared in the sentence, the corresponding vector location is set to 1, all other locations are set to 0. This feature extraction allowed us to preserve the information of specific vocabulary in each sentence. Similarly with the method one, a total of 4000 sentences were used to train the SVM model, and 500 sentences were used to evaluate the model. The same testing dataset was used, and the model is able to successfully classify the test data into two classes. With further inspection into the dataset, we found out that the majority data for these two classes contains unique words, and for other classes, there are no specific matching words for the classes. This characteristic of dataset explained the behavior of this SVM classifier. The learning rate and overall performance of the model is shown in the figure below

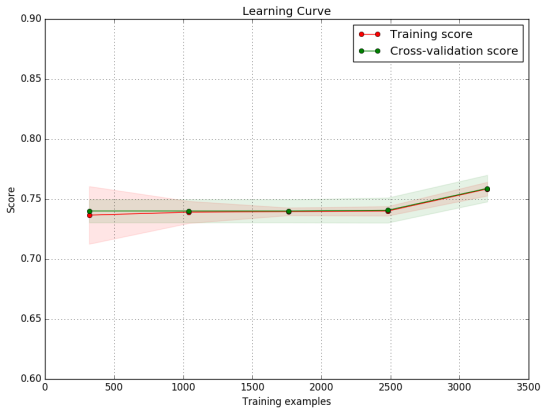

Figure 2: SVM Learning rate with second feature extraction

For the recurrent neural net with linear regression, we convert sentence to 15 integers as we did in the first sentence extraction approach.These features was treated as the input of a basic RNN cell. The RNN cell from tensorflow was being used to generate 40 features with time step 15. Meanwhile, a dictionary of label was build on all labels of ATIS dataset. This dictionary was converted to one hot vector of length 17. The features from last output of RNN cell was trained by a linear regression and generated 17 features. A softmax was applied to these features to generate scores. A gradient decent optimizers was applied to minimize cross entropy of scores and labels. Similarly, we trained 4000 examples and use 500 sentences as test set. The classification accuracy of training set is 0.758, and accuracy for test set is 0.732.

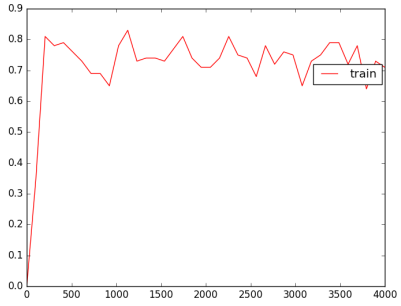

Figure 3: RNN with linear regression

5 Next Step

For the next step, we are going to try to modify our construction of features for SVM. The current features are based on only the non-semantic characteristics of the sentence. If we could extract features with more appropriate methods like word embedding, the performance of the classification of SVM could be better. Besides, because we just used the ATIS dataset for the training and testing, we are also going to try different dataset for general intent classification. e.g. social media dataset. Then we are going to focus on using Recurrent Neural Network(RNN) to extract the intent. As we mentioned, we just began to build a draft model to use RNN to do the classification. So we plan to put more efforts on the feature extraction for RNN. Besides, based on the feature of the spoken language text, we are going to use Long Short Term Memory cell to construct the RNN in the next move. Besides, we may try to use RNN to perform slot filling task by using word embedding if possible. As we could get more accurate understanding of the sentence if we have the slot filling label for each word.

6 Reading List

1. Contextual Spoken Language Understanding Using Recurrent Neural Networks [2]
2. Convolutional Neural Network Based Triangular CRF for Joint Intent Detection and Slot Filling [5]
3. Liblinear: A library for large linear classification [1]
4. Contextual domain classification in spoken language understanding systems using recurrent nerual network [4]
5. Understanding Spoken Language [3]

References

[1] Ran R., Chang K., Hsieh C., Wang X., Lin C. “Liblinear: A library for large linear classification”. In: *Journal of MAchine Learning Research* (2008).

[2] Shi Y., Yao K., Chen H., Pan Y.C., Hwang M.Y., Peng B. “Contextual Spoken Language Understanding Using Recurrent Neural Networks”. In: (2015).

[3] Tur G., Wang, Y., Hakkani-Tr D. “Understanding Spoken Language”. In: *Chapman and HallCRC Press* (2013).

[4] Xu P., Sarikaya R. “Contextual domain classification in spoken language understanding systems using recurrent nerual network”. In: *ICASSP* (2014).

[5] Xu P., Sarikaya R. “Convolutional Neural Network Based Triangular CRF for Joint Intent Detection and Slot Filling”. In: (2013).