

Title	MLLRにおける回帰行列の重み付き線形和を用いた適応法に関する研究
Author(s)	小山, 岳史
Citation	
Issue Date	2005-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1916
Rights	
Description	Supervisor:党 建武, 情報科学研究科, 修士

A study on speaker adaptation based on MLLR using linear combination of regression matrices

Takefumi Koyama (210037)

School of Information Science,
Japan Advanced Institute of Science and Technology

February 10, 2005

Keywords: HMM, speaker adaptation, MLLR, combination of regression matrices.

1 Introduction

In speech recognition that uses HMM, generally, as for recognition that uses the unspecified speaker model, the recognition rate is lower than the case to use the specific speaker model. It is thought that because the parameter of HMM used for recognition doesn't suit the speaker. Then, we have to improve the recognition performance by adjusting the parameter of HMM by using speaker's voice sample data. This method is Speaker Adaptation.

MLLR(Maximum Likelihood Linear Regression) is a general speaker adaptation method. In MLLR, the acoustic model is divided into some regression classes, and regression matrices are calculated in each class. By this method, it is possible to adapt even if the model without the adaptation data. However, it is thought that you should do a different adaptation a model in the vicinity of the center of the class and a model away from the center even if they are in the same class.

In this research, we propose the technique for calculating the regression matrices suitable the model by using linear combination of the regression matrices.

2 Adaptation using linear combination of regression matrices

The method based on the bhattacharyya distance was used as a computational method of the coefficient of the matrices. A new regression matrix W_{new} of model is calculated by using d_i and W_i . (d_i is distance to center of class C_i , and W_i is regression matrix of C_i . N is number of regression classes.)

$$W_{new} = c_1W_1 + c_2W_2 + \cdots + c_NW_N$$

where

$$c_i = \frac{D_i}{\sum_{k=1}^N D_k}, \quad D_i = \frac{1}{d_i}$$

By the value of d_i small, the value of c_i grows. $\sum_{k=1}^N c_i$ is 1.

3 Selection the model which used the proposal method

To take the effect of the adaptation or more, the proposal method was used only for a model away from the center of the class to some degree. We choosed a model that other classes are smaller than the distance to the center of the class to which the model belongs.

4 Selection the matrices which used the proposal method

As we calculate linear combination of regression matrices, we thought that we did not have to use the matrix of a class away from the model to some degree. In this research. We use only matrices of classes near from the model to some degree. As the standard, the distance a model and center of class.

d_i is the distance to center of class C_i and the model in C_i . d_j is the distance to center of class C_j . if

$$d_j \geq d_i$$

then, we use the matrix W_j which regression matrix of C_j .

5 Experiment of speaker adaptation

As a result of the experiment that compared the proposal methods with MLLR, the proposal method was slightly effective when the number of adaptation words was 10-20. But when the number of adaptation words was more than 30, the recognition rate fell more than MLLR.

When there are comparatively a lot of numbers of adaptation words, the regression class is decided robustly. Therefore, it is thought that it is because adding the hand in this state becomes a contrary effect.

However, by additionally using selection models and matrices, the recognition rate of the proposal method was slightly exceeded or the same level as MLLR.

6 Conclusion and subject of feature investigation

The proposal method showed the possibility of the improvement of the performance concerning the model around the regression class though the effect in improvement of the recognition rate was a little.

Subject of feature investigation is an examination of the calculation method of the coefficient of the matrix, selection the best models and using matrices.