Studentská Vědecká Konference 2010

SMOOTHING FACTOR IN DISCRIMINATIVE FEATURE ADAPTATION

Zbyněk ZAJÍC [1]

1 INTRODUCTION

In these days, Discriminative Training (DT) methods of an acoustics model are taking over the leadership in the speaker recognition task for training an acoustics model. Maximum Likelihood (ML) training suffers from some inaccuracies because of improper assumptions of the suitability of the HMM. Well-known adaptation method, feature Maximum Likelihood Linear Regression (fMLLR), is based on ML criterion:

$$\mathcal{F}_{ML}(\boldsymbol{\lambda}) = p(\boldsymbol{O}|W_{ref}, \boldsymbol{\lambda}), \tag{1}$$

where $\boldsymbol{O}$ represents the sequence of feature vectors related to one speaker, W_{ref} is the corresponding correct transcription and $\boldsymbol{\lambda}$ denotes the set of Hidden Markov Model (HMM) parameters. ML criterion is optimized by Expectation-Maximization (EM) algorithm, but this approach has two limitation Yu (2006). The first is the assumption that training data bring a good generalization for testing data. The second is the amount of data to train a large, complex model. Both limitations are usually hard to satisfy.

While the ML criterion try to maximize the likelihood of the observation states sequence given the correct transcriptions, the DT criteria reflect the recognition error and try to minimized it. Maximum Mutual Information (MMI) criterion in Yu (2006) is one of the DT possibilities

$$\mathcal{F}_{MMI}(\boldsymbol{\lambda}) = \frac{p(\boldsymbol{O}|W_{ref}, \boldsymbol{\lambda})P(W_{ref})}{\sum_W p(\boldsymbol{O}|W, \boldsymbol{\lambda})P(W)}, \tag{2}$$

where W is a transcription with all possible hypothesis. MMI increases the posterior probability of model states corresponding to their adaptation data (numerator in (2), similar with (1)) and decreases the probability of confusion data (denominator in (2)) at the same time.

The main problem consists in the optimization process, where mainly the weak-sense auxiliary function is used. Regrettably, it does not guarantee the convergence of the discriminative criterion. In order to adjust the stability of discriminative criteria a smoothing term is involved. Another criteria are e.g. Minimum Phone Error (MPE) or Minimum Classification Error (MCE).

2 DISCRIMINATIVE FEATURE MAXIMUM LIKELIHOOD LINEAR REGRESSION (DFMLLR)

DfMLLR technique belongs to the category of Discriminative Linear Transformations (DLTs) and like its non-discriminative version fMLLR described in Povey (2006), DfMLLR transforms feature $\boldsymbol{o}_t$ according to

$$\bar{\boldsymbol{o}}_t = \boldsymbol{A}\boldsymbol{o}_t + \boldsymbol{b}, \tag{3}$$

[1]Ing. Zbyněk Zajíc, Ph.D. student,University of West Bohemia in Pilsen, Faculty of Applied Sciences, Department of Cybernetics, Univerzitní 22, 306 14 Pilsen, e-mail: zzajic@kky.zcu.cz

The estimation formulas for transformation matrices $\boldsymbol{A}$ and $\boldsymbol{b}$ can be found in Wang (2004).

DfMLLR does not access the data directly, but only through accumulated statistics (formulas can be found in Zajíc (2009)), which are cumulated in the first step of the adaptation process. These statistics are $jm-th$ mixtures' posterior $\gamma_{jm}(t)$, its sum for all adaptation features and sum of the first and the second moment of features aligned to the $jm-th$ mixture . For discriminative approach also denominator statistics for confusable states must be accumulated. These are computed in the sense of the denominator in (2).

As mentioned in the introduction, the primary weakness of discriminative methods is the need to utilize weak-sense auxiliary function in order to find the solution of the criterion. In DfMLLR, MMI estimation of transformations is confronted with ML estimation (through smoothing factor) to avoid the instability. In Wang (2003) the smoothing factor depends on the estimated mean of adapted data. In adaptation, there is usually no sufficient amount of data. Another solution proposed by Wang (2004) is involvement of the fMLLR-adapted mean vector, which is more time-consumed.

To solve this problem another two alternatives are introduced. When the smoothing factor is computed, the original mean vector can be used. The advantage of this approach is its speed. Another possibility is using Maximum A-posteriori Probability (MAP) estimation of mean vector (Gauvain (1994)), which is faster then fMLLR estimation. As can be seen from results, all methods have similar accuracy, but the use of the original mean vector does not involve any additional computation.

3 CONCLUSION

Discriminative criteria, especially MMI criterion, were introduced. These criteria are suitably utilized in the adaptation process and bring a significant improvement in comparison to non-discriminative ones (1.3% relativelly in Zajíc (2009)). The requirement of the smoothing factor is caused by the use of the weak-sense auxiliary function. Two different approaches for smoothing factor were proposed in this paper, MAP estimation or original mean vector. Solution for the smoothing factor defined in the literature and proposed in this paper was found consistent in the sense of the efficiency, but the proposed solution is less time-consumming.

Acknowledgement: The work has been supported by the grant of The University of West Bohemia, project No. SGS-2010-054. The access to the MetaCentrum clusters provided under the research intent MSM6383917201 is appreciated.

REFERENCES

Kai Yu, 2006. Adaptive Training for Large Vocabulary Continuous Speech Recognition. PhD Thesis, Cambridge University.

Zajíc, Z., Machlica, L., Müller, L., 2009. Refinement approach for adaptation based on combination of MAP and fMLLR. In: TSD, pp. 274-281, Pilsen.

Wang, L., Woodland P.C., 2004. MPE-based discriminative linear transformation for speaker adaptation. In: IEEE International Conference on ASSP, pp. 321-324.

Wang, L., Woodland P.C., 2003. Discriminative adaptive training using the MPE criterion. In: IEEE Automatic Speech Recognition and Understanding, pp. 279-284.

Povey, D., Saon, G., 2006. Feature and Model Space Speaker Adaptation with Full Covariance Gaussians. In: Interspeech, paper 2050-Tue2BuP.14.

Gauvain, L., Lee, C.H., 1994. Maximum A-Posteriori Estimation for Multivariate Gaussian Mixture Observations of Markov Chains. In: IEEE Transactions SAP, pp.291–298.