Computer Speech and Language (2001) **15**, 15–38
doi:10.1006/csla.2000.0151

Discounted likelihood linear regression for rapid speaker adaptation

Asela Gunawardana and William Byrne

Center for Language and Speech Processing, Department of Electrical and Computer Engineering, The Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218, U.S.A.

Abstract

The widely used maximum likelihood linear regression speaker adaptation procedure suffers from overtraining when used for rapid adaptation tasks in which the amount of adaptation data is severely limited. This is a well known difficulty associated with the expectation maximization algorithm. We use an information geometric analysis of the expectation maximization algorithm as an alternating minimization of a Kullback–Leibler-type divergence to see the cause of this difficulty, and propose a more robust discounted likelihood estimation procedure. This gives rise to a discounted likelihood linear regression procedure, which is a variant of maximum likelihood linear regression suited for small adaptation sets. Our procedure is evaluated on an unsupervised rapid adaptation task defined on the Switchboard conversational telephone speech corpus, where our proposed procedure improves word error rate by 1.6% (absolute) with as little as 5 seconds of adaptation data, which is a situation in which maximum likelihood linear regression overtrains in the first iteration of adaptation. We compare several realizations of discounted likelihood linear regression with maximum likelihood linear regression and other simple maximum likelihood linear regression variants, and discuss issues that arise in implementing our discounted likelihood procedures.

1. Introduction

Acoustic models used in speaker-independent (SI) large vocabulary speech recognition are trained using data from a large number of speakers. Speaker adaptation procedures attempt to reestimate the acoustic models so that they better match speakers encountered during recognition. Since the amount of acoustic data available for this is typically small, the reestimation of the acoustic models needs to be done in a constrained manner, to ensure that the resulting adapted acoustic models are robust. In the case of rapid adaptation, where adaptation data are extremely scarce, this robustness issue is especially problematic.

There are two common approaches to dealing with this problem of data scarcity. The first consists of choosing the adapted models from a small set of constrained acoustic models which are anchored at the SI acoustic models. The maximum likelihood linear regression (MLLR) scheme (Digalakis, Rtischev & Neumeyer, 1995; Leggetter & Woodland, 1995) discussed in the next section is an example of such a scheme. The second approach is to

0885–2308/01/010015 + 24 $35.00/0

use an estimation procedure that is more robust than the standard expectation-maximization (EM)-based maximum likelihood procedure. The maximum *a posteriori* (MAP) adaptation procedure (Lee & Gauvain, 1993) is an example of this approach—a prior distribution is used to penalize the selection of acoustic models that differ greatly from the speaker-independent acoustic models.

The use of a prior on the MLLR parameters themselves (Chou, 1999; Wang & Zhao, 2000) is an example of a combination of these approaches—the model constraints of MLLR are combined with a MAP estimation procedure. Our proposed adaptation scheme is also such an approach. We propose a general robust *discounted likelihood* estimation procedure and use it for adaptation under the same constraints as MLLR. However, unlike the MAP procedure, our approach does not require a prior distribution to constrain the estimation.

In this paper, we address the rapid adaptation of acoustic hidden Markov models (HMMs) with Gaussian mixture output distributions. In this case, it is convenient to think of estimating adapted acoustic HMMs as estimating transformations of the SI acoustic HMM parameters. In the case of MLLR, these transformations are constrained to be affine transformations of the Gaussian means. These transformations are estimated from adaptation data using the EM algorithm (Dempster, Laird & Rubin, 1977). Unfortunately, the EM algorithm gives unreliable estimates when a small amount of data is used. This behavior can be understood using the information geometric description of EM which was presented by Csiszár and Tusnády (1984), and applied to acoustic model estimation by Ephraim and Rabiner (1990). Using this description, we will see that MLLR finds transformations that put as much probability mass as possible on the adaptation data. However, since we have only a small amount of adaptation data which may be unrepresentative of the acoustics of the test speaker, this is not desirable, and leads to the well known overtraining property of the maximum likelihood estimator. We use the insight gained from this description to formulate a new, more robust estimation technique which we call *discounted likelihood estimation*, and apply it to the problem of estimating MLLR-type affine transformations of the Gaussian means, yielding an adaptation scheme which we term *discounted likelihood linear regression* (DLLR).

Before presenting the discounted likelihood estimation scheme, we briefly review MLLR in Section 2. In Section 3, we present an analysis of how the EM algorithm which is used in MLLR estimation can be viewed as an alternating minimization procedure, and in doing so, we will see why MLLR is not robust when used for rapid adaptation. This motivates the presentation of the discounted likelihood estimation procedure in Section 4, and we derive reestimation equations for adaptation in Section 5. In Section 6, we discuss certain modeling issues which arise in the practical application of these reestimation equations, and then summarize the algorithms implemented. In Section 7 we discuss how the DLLR schemes presented relate to other estimation techniques. We evaluate our DLLR schemes in Section 8 and conclude in Section 9. In Table I, we summarize notation that will be used throughout the paper.

2. Speaker adaptation with MLLR

The popular maximum likelihood linear regression adaptation scheme parameterizes the Gaussian emission distributions of the acoustic HMMs by affine transformations of the SI means. In this paper, we follow Leggetter and Woodland (1995) and discuss only the single Gaussian output density case, and note that the multiple Gaussian mixture case is a straight-