Return to Main

Objectives

Introduction:
Motivation

Approaches:
MAP
MLLR
Comparison

On-Line Resources:
MAP
MLLR
Comparison

LECTURE 39: ADAPTATION

- Objectives:
  - Maximum a posteriori estimation
  - Maximum likelihood linear regression
  - Comparison in performance

This lecture draws on material from the course textbook:

> X. Huang, A. Acero, and H.W. Hon, *Spoken Language Processing - A Guide to Theory, Algorithm, and System Development*, Prentice Hall, Upper Saddle River, New Jersey, USA, ISBN: 0-13-022616-5, 2001.

and this presentation/paper:

> J. Hamaker, "A Speaker Adaptation Techniques for LVCSR", http://www.isip.msstate.edu/publications/courses/ece_7000_speech/lectures/current/lecture_10/, ECE 7000: Special Topics in Speech Recognition, Department of Electrical and Computer Engineering, Mississippi State University, Mississippi, USA, November 1999.

LECTURE 39: ADAPTATION

- Objectives:
  - Maximum a posteriori estimation
  - Maximum likelihood linear regression
  - Comparison in performance

This lecture draws on material from the course textbook:

X. Huang, A. Acero, and H.W. Hon, *Spoken Language Processing - A Guide to Theory, Algorithm, and System Development*, Prentice Hall, Upper Saddle River, New Jersey, USA, ISBN: 0-13-022616-5, 2001.

and this presentation/paper:

J. Hamaker, "A Speaker Adaptation Techniques for LVCSR", http://www.isip.msstate.edu/publications/courses/ece_7000_speech/lectures/current/lecture_10/, ECE 7000: Special Topics in Speech Recognition, Department of Electrical and Computer Engineering, Mississippi State University, Mississippi, USA, November 1999.

ADAPTIVE TECHNIQUES - MINIMIZING MISMATCH

- We can improve recognition performance by training on a single speaker. This is known as *speaker dependent* speech recognition.
- However, there are numerous training problems (long enrollment). An alternate approach is to *adapt* speaker independent models.
- Such adaptation techniques are generally used to reduce mismatch between the acoustic models and the decoding environment (e.g., microphone, acoustic channel and speaker mismatch).
- There are two basic approaches:
    - **Maximum A Posteriori (MAP)**: choosing an estimate that maximizes the posterior probability (consistent with the observed data and prior inormation).
    - **Maximum Likelihood Linear Regression (MLLR)**: ML estimate of a linear transformation.

MAXIMUM A POSTERIORI ADAPTATION

Given observation data $\boldsymbol{X}$, the MAP estimate is:

$$\hat{\Phi} = \arg\max_{\Phi} [p(\boldsymbol{X}|\Phi)p(\Phi)]$$

If we have no prior information, $p(\Phi)$ is the uniform distribution, and the MAP estimate is identical to the MLE estimate. However, if we have prior information, we can use EM to estimate the new parameters:

$$Q_{MAP}(\Phi,\hat{\Phi}) = \log p(\Phi) + Q(\Phi,\hat{\Phi})$$

Under a significant number of assumptions, we can derive a rather simple and intuitive expression for updating Gaussian means:

$$\hat{\mu}_{ik} = \frac{c_{ik}}{\tau_{ik} + c_{ik}}\mu'_{ik} + \frac{c_{ik}}{\tau_k + c_{ik}}\mu_{ik}$$

where:

μ'_{ik}: ML estimate of the mean

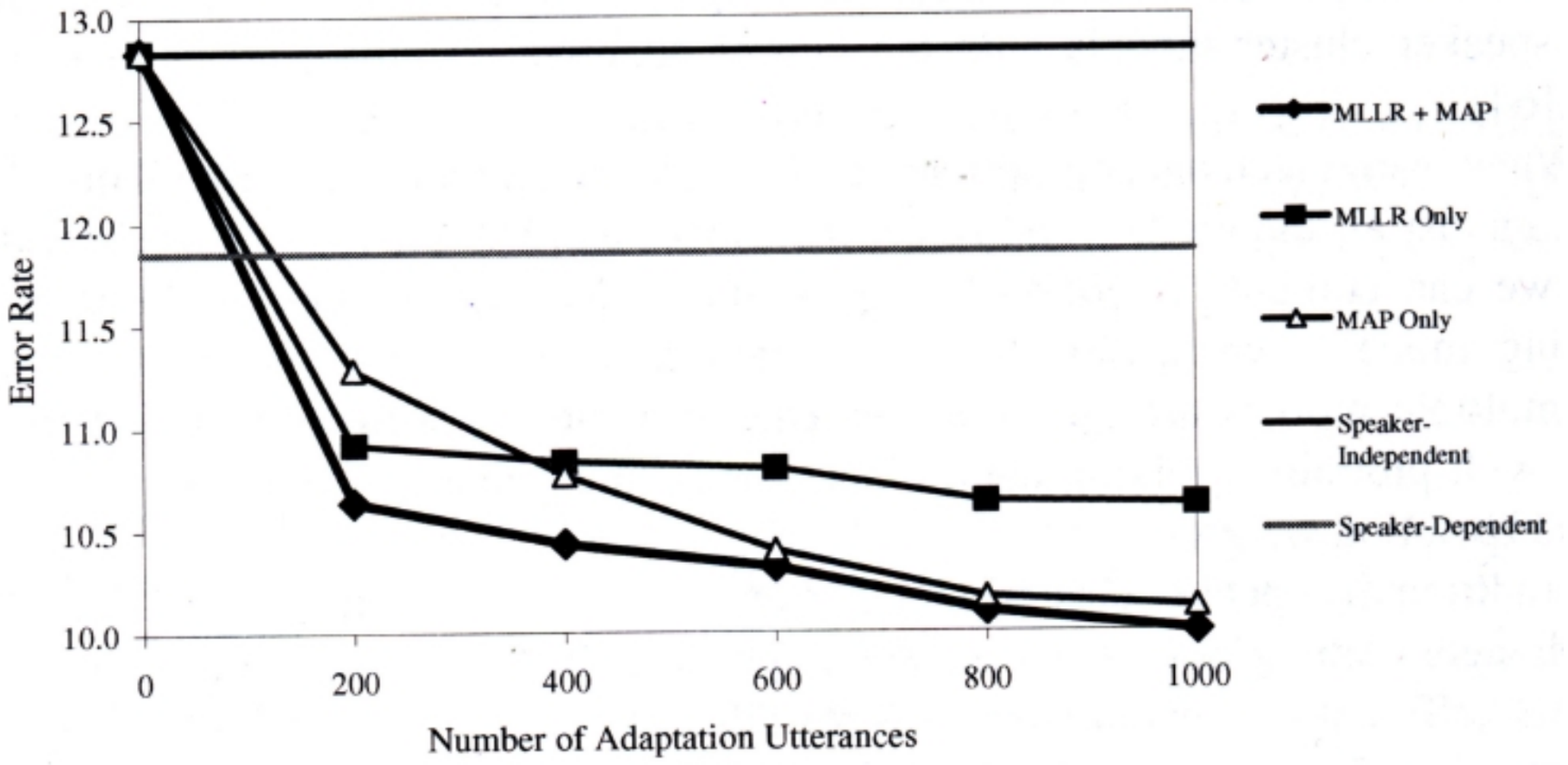
μ_{ik}: existing estimate of the mean (prior information)

c_{ik}: $\sum_{t=1}^{T} \zeta_{ik}(i, k)$

τ_{ik}: parameter controlling relative weight of prior information

A COMPARISON OF MLLR AND MAP

Below is a comparison of MLLR and MAP on a 60,000 word dictation task. The speaker dependent system was trained on 1,000 sentences.



Though MAP appears to be fairly powerful in this example, MLLR is much more popular. MLLR+MAP combines the best of both approaches, but also leads to a more complicated system.

MLLR: A SPEAKER ADAPTATION TECHNIQUE FOR LVCSR

Jon Hamaker
Institute for Signal and Information Processing
Mississippi State University, Mississippi State, MS 39762
email: hamaker@isip.msstate.edu

ABSTRACT

In typical state-of-the-art large vocabulary conversational speech recognition (LVCSR) systems a single model is developed using data from a large number of speakers to cover the variance across dialects, speaking styles, etc. With this, we expect that our systems will generalize well to any particular speaker. However, from experience we know that there are speakers who are poorly modeled using this paradigm. Thus, it would be advantageous to adapt the models, during run-time, to the new speaker. Following this premise, many methods have been developed which use a small amount of a speaker's data to adapt the speaker-independent model to a speaker-dependent one.

In this talk we will review the motivation and methodology behind these methods. Much of the time will be spent in describing one popular method which uses a maximum likelihood linear regression (MLLR) approach to speaker adaptation. MLLR builds a transform for the model parameters using linear regression so that the transformed parameters of each model better represent the new speaker. Applying this approach to all of the models in an LVCSR system (particularly when using mixture models) would require an unreasonable number of additional parameters and a large amount of training data for full coverage. To attack this problem a small number of transforms are built and tying is used. MLLR has become a standard feature in most LVCSR systems and has proven successful in every major speaker-independent speech recognition task to which it has been applied.

Additional items of interest:

- Presentation
- Paper
- Software, Data, etc.

MLLR: A SPEAKER ADAPTATION TECHNIQUE FOR LVCSR

Jon Hamaker
Institute for Signal and Information Processing
Mississippi State University, Mississippi State, MS 39762
email: hamaker@isip.msstate.edu

ABSTRACT

In typical state-of-the-art large vocabulary conversational speech recognition (LVCSR) systems a single model is developed using data from a large number of speakers to cover the variance across dialects, speaking styles, etc. With this, we expect that our systems will generalize well to any particular speaker. However, from experience we know that there are speakers who are poorly modeled using this paradigm. Thus, it would be advantageous to adapt the models, during run-time, to the new speaker. Following this premise, many methods have been developed which use a small amount of a speaker's data to adapt the speaker-independent model to a speaker-dependent one.

In this talk we will review the motivation and methodology behind these methods. Much of the time will be spent in describing one popular method which uses a maximum likelihood linear regression (MLLR) approach to speaker adaptation. MLLR builds a transform for the model parameters using linear regression so that the transformed parameters of each model better represent the new speaker. Applying this approach to all of the models in an LVCSR system (particularly when using mixture models) would require an unreasonable number of additional parameters and a large amount of training data for full coverage. To attack this problem a small number of transforms are built and tying is used. MLLR has become a standard feature in most LVCSR systems and has proven successful in every major speaker-independent speech recognition task to which it has been applied.

Additional items of interest:

- Presentation
- Paper
- Software, Data, etc.