ELSEVIER

Speech Communication 37 (2002) 321–334

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

A Bayesian prediction approach to robust speech recognition and online environmental learning

Jen-Tzung Chien *

Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan 70101, Taiwan, ROC

Received 7 August 2000; received in revised form 27 February 2001; accepted 28 February 2001

Abstract

A robust speech recognizer is developed to tackle the inevitable mismatch between training and testing environments. Because the realistic environments are *uncertain* and *nonstationary*, it is necessary to characterize the uncertainty of speech hidden Markov models (HMMs) for recognition and trace the uncertainty incrementally to catch the newest environmental statistics. In this paper, we develop a new Bayesian predictive classification (BPC) for *robust decision* and *online environmental learning*. The BPC decision is adequately established by modeling the uncertainties of both the *HMM mean vector* and *precision matrix* using a conjugate prior density. The *frame-based predictive distributions* using multivariate t distributions and approximate Gaussian distributions are herein exploited. After the recognition, the prior density is pooled with the likelihood of the current test sentence to generate the reproducible prior density. The hyperparameters of the prior density are accordingly adjusted to meet the newest environments and apply for the recognition of upcoming data. As a result, an *efficient online unsupervised learning* strategy is developed for HMM-based speech recognition without needing adaptation data. In the experiments, the proposed approach is significantly better than conventional plug-in maximum a posteriori (MAP) decision on the recognition of connected Chinese digits in hands-free car environments. This approach is economical in computation. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Bayesian predictive classification (BPC); Online unsupervised learning; Speaker adaptation; Speech recognition; Hidden Markov model

1. Introduction

There is no doubt that the robustness issue is increasingly important in the community of speech recognition because the mismatches between training and testing data always exist in real-world applications and seriously degrades the recognition performance (Lee, 1998; Huo, 1998; Lee and Huo, 1999). Basically, the mismatches come from the variabilities of inter- and intra-speakers, transducers/channels and surrounding noises. For the applications of hands-free speech recognition in car environments, the recognizer needs to handle the low signal-to-noise ratio (SNR) speech signal caused by the far-talking microphone and the ambient noises such as engine, babble, wind, shower, music, broadcast, etc., under different driving

* Tel.: +886-6-2757575 ext. 62532; fax: +886-6-2747076.
E-mail address: jtchien@mail.ncku.edu.tw (J.-T. Chien).

0167-6393/02/$ - see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0167-6393(01)00032-2