Discriminative and Bayesian techniques for hidden Markov model speech recognition systems

by

Darryl William Purnell

Submitted in partial fulfillment of the requirements for the degree of

PHILOSOPHIAE DOCTOR (ENGINEERING)

in the Faculty of Engineering

UNIVERSITY OF PRETORIA

February 2001

Summary

The collection of large speech databases is not a trivial task (if done properly). It is not always possible to collect, segment and annotate large databases for every task or language. It is also often the case that there are imbalances in the databases, as a result of little data being available for a specific subset of individuals. An example of one such imbalance is the fact that there are often more male speakers than female speakers (or *vice-versa*). If there are, for example, far fewer female speakers than male speakers, then the recognizers will tend to work poorly for female speakers (as compared to performance for male speakers).

This thesis focuses on using Bayesian and discriminative training algorithms to improve continuous speech recognition systems in scenarios where there is a limited amount of training data available. The research reported in this thesis can be divided into three categories:

- Overspecialization is characterized by good recognition performance for the data used during training, but poor recognition performance for independent testing data. This is a problem when too little data is available for training purposes. Methods of reducing overspecialization in the minimum classification error algorithm are therefore investigated.
- Development of new Bayesian and discriminative adaptation/training techniques that can be used in situations where there is a small amount of data available.

One example here is the situation where an imbalance in terms of numbers of male and female speakers exists and these techniques can be used to improve recognition performance for female speakers, while not decreasing recognition performance for the male speakers.

- Bayesian learning, where Bayesian training is used to improve recognition performance in situations where one can only use the limited training data available. These methods are extremely computationally expensive, but are justified by the improved recognition rates for certain tasks. This is, to the author's knowledge, the first time that Bayesian learning using Markov chain Monte Carlo methods have been used in hidden Markov model speech recognition.

The algorithms proposed and reviewed are tested using three different datasets (TIMIT, TIDIGITS and SUNSpeech), with the tasks being connected digit recognition and continuous speech recognition. Results indicate that the proposed algorithms improve recognition performance significantly for situations where little training data is available.

Keywords: speech recognition, hidden Markov model training, minimum classification error, Bayesian adaptation, Bayesian learning, maximum a posteriori parameter estimation, sparse data

Samevatting

Die versameling van groot spraakdatabasisse is nie 'n maklike taak nie. Dit is nie altyd moontlik om groot databasise to versamel, in segmente te verdeel en te annoteer vir elke taak of taal nie. Dit is ook dikwels die geval dat daar 'n wanbalans bestaan in 'n databasis, as gevolg van die onverkrygbaarheid van data vir 'n spesifieke subgroep van individue. Een voorbeeld van so 'n wanbalans is die feit dat daar dikwels meer manssprekers as vrouesprekers is (of andersom). In so 'n geval, sal die herkenner gewoonlik nie goed werk vir vrouens nie, maar sal relatief goed werk vir mans.

Hierdie tesis fokus op die gebruik van Bayes en diskriminerende afrigtingstegnieke om kontinuspraakherkenningstelsels te verbeter in scenarios waar min afrigdata beskikbaar is. Die navorsing waaroor hier gerapporteer word, kan in drie dele verdeel word:

- Oor-spesialisasie word gekarakteriseer deur goeie herkenning vir die afrigdata, maar slegte herkenning vir onafhanklike toetsdata. Hierdie probleem onstaan wanneer te min data beskikbaar is vir afrigtingsdoeleindes. Metodes om oor-spesialisasie in minimum klassifikasiefout afrigting te verminder word dus hier ondersoek.

- Nuwe Bayes en diskriminerende afrigtings- en aanpassingstegnieke om herkenning te verbeter in situasies waar min data beskikbaar is. Een voorbeeld is die situasie waar 'n wanbalans in terme van die aantal vroue- en manssprekers bestaan. Hierdie tegnieke kan gebruik work om die herkenningstempo te verbeter vir die

vrouesprekers.

- Bayes afrigting word gebruik om herkenning te verbeter in situasies waar min data beskikbaar is, en geen taakspesifieke data nie. Hierde metode is uiters berekeningintensief, maar ek glo dat die verbetering in herkenningstempo dit regverdig. Hierdie is, na ons medewete , die eerste keer dat Markov ketting Monte Carlo gebaseerde Bayesian afrigting in verskuilde Markov model spraakherkenningstelsels gebruik word.

Die voorgestelde algoritmes is getoets met drie verskillende spraakdatabasise: TIMIT, TIDIGITS en SUNSpeech. Die take is kontinusyferherkenning en kontinuspraakherkenning. Resultate toon dat die voorgestelde algoritmes goed werk vir situasies waar min afrigtingsdata beskikbaar is.

Sleutelwoorde: spraakherkenning, verskuilde Markov model afrigting, minimum klassifikasiefout, Bayes aanpasing, Bayes afrigting, maksimum a posteriori parameter skating, min afrigdata

Acknowledgements

First and foremost, I would like to thank Professor Liesbeth Botha for the advice and guidance she gave me during the last three years.

I would also like to thank Christoph Nieuwoudt and Johann Holm, with whom I had many interesting and thought-provoking discussions.

Finally, I would like to thank my family and friends for their support and encouragement throughout.

Contents