A ROBUST TRAINING STRATEGY AGAINST EXTRANEOUS ACOUSTIC VARIATIONS FOR SPONTANEOUS SPEECH RECOGNITION

Hui Jiang† *and Li Deng*‡

Department of Electrical and Computer Engineering, University of Waterloo, Canada
† Currently Multimedia Communications Research Lab, Bell Labs, Murray Hill, NJ 07974
‡ Currently Microsoft Research, Redmond, WA 98052
Email: hui@research.bell-labs.com and deng@microsoft.com

ABSTRACT

In the paper, we propose a robust training strategy to deal with extraneous acoustic variations for conversational speech recognition. This strategy generalizes speaker adaptive training, where HMM parameter transformations are used to normalize the extraneous variations in the training data according to a set of pre-defined *conditions*. Then a compact model and the associated prior p.d.f.'s of transformation parameters are estimated using the maximum likelihood criterion. In the testing phase, the compact model and the prior p.d.f.'s are used to search for the unknown word sequence based on Bayesian Prediction Classification. The proposed strategy is evaluated in a Switchboard task to deal with pronunciation variations in spontaneous speech recognition. Preliminary results show moderate word error rate reduction over a well-trained baseline system under identical experimental conditions.

1. INTRODUCTION

In the past few decades, the statistical models, such as hidden Markov models (HMM), have achieved significant success in automatic speech recognition (ASR). In the conventional paradigm of ASR, the statistical models are usually estimated from a huge amount of training data. The training data usually are collected under as many different conditions as possible for the purpose of properly representing possible incoming speech data in the future use. Even though the data collection conditions greatly may differ due to a wide range of factors, the conventional paradigm treats all train data collected in different conditions in an identical manner by simply pooling them together. Then the model parameters are resolved from the pooled data set via some parameter estimation techniques, e.g. ML criterion and discriminant training. An apparent shortcoming of this training paradigm is that the large amount of pooled training data not only include the relevant variability (such as phonetic distinction), but also involve many other extraneous variations which are irrelevant to our modeling purpose and should therefore be compensated for. We call the variations in the data which have nothing to do with our modeling purpose as *extraneous variations*. For instance, in a typical case of speech recognition, it is only necessary to model the phonetically relevant variation sources. All other variabilities are considered to be extraneous, including those arising from speaker, transducer, telephone channel, speaking style, speaking rate, pronunciation change, etc.

In the conventional implementation of speech recognizers, we do not have an explicit mechanism to compensate these extraneous and irrelevant variations in the training procedure. In particular, when we recognize spontaneous speech where many types of extraneous variations abound, speech recognition performance can be significantly affected. In the training phase, due to the extraneous variations, training data may deviate from what is assumed in the model. This would make the estimated models diverge from the desired behavior. In the testing phase, the deviation due to the extraneous variation can also be viewed as a special kind of mismatch between the models and the testing data. In this paper, we describe a robust strategy to deal with the extraneous acoustic variations in the training phase only. How to handle them in the testing phase is currently under investigation.

Recently, some researchers began to notice the importance to compensate the extraneous variations in the training phase in order to improve the generalization capability of the models. In [1], the "speaker-adaptive training" (SAT) by BBN researchers is one of important steps along this direction. The work reported in [4] shows another way to normalize irrelevant variability in the training phase, but for the purpose of learning a model structure (HMM state tying). In this paper, we propose a new robust training strategy to compensate and/or normalize the extraneous variations, with more elegant theoretical foundation and practical effectiveness. Briefly, we label each utterance in the training set with a pre-defined *condition*, which could depend on speaker *id*, speaking style, pronunciation, transducer, transmission channel, etc. The data from different conditions are first normalized by using appropriate transformations before they are pooled together to estimate a "compact model". Meanwhile, a prior distribution of transformation parameters is estimated to represent the knowledge of all possible transformations used in the training phase across the "conditions". In this way, the extraneous variation is adequately compensated for and the compact model can then converge properly to represent the relevant variations in question. In the testing/decoding phase, based on the compact model and the prior distribution of transformation parameters, we use a new search algorithm to decode any new input utterance according to Bayesian prediction[6].

In this work, the proposed strategy is used to normalize and/or compensate the extraneous variations in the Switchboard task in order to obtain better acoustic models which can adequately describe phonetically relevant variation sources. According to [2, 8], in the Switchboard task, pronunciation variations in conversational speech is one major extraneous variation source hampering speech recognition. Therefore, we have in this work specifically focused on the "condition" that characterizes the pronunciation variation. To facilitate the implementation, we have also chosen very sim-

ple transformations, i.e., piecewise linear functions, to normalize and/or compensate pronunciation variations of conversational speech in the Switchboard task.

2. OVERVIEW OF THE NEW STRATEGY

Following the idea originally presented in [1], suppose we have a compact HMM model $\lambda_c = \{\pi_i, a_{ij}, w_{ik}, m_{ik}, r_{ik} \mid 1 \le i, j \le N, 1 \le k \le K\}$ for each speech unit W we desire to model, and all training data for W is composed of $X = \{X^{(r)} \mid r = 1, 2, \cdots, R\}$, where $X^{(r)}$ denotes those data collected under the condition r. Here each condition corresponds to a distinct pronunciation of the word W. In each state of the compact model, we have the state distribution of HMM with diagonal precision matrix as

$$p_i(x) = \sum_{k=1}^{K} w_{ik} \cdot \prod_{d=1}^{D} \sqrt{\frac{r_{ikd}}{2\pi}} \exp[-\frac{r_{ikd}}{2}(x_d - m_{ikd})^2] \quad (1)$$

where D denotes the dimension of feature vectors.

Here we aim to choose some proper transformations to normalize/compensate the pronunciation variations in conversational speech. In other words, we need to choose a set of transformations for model λ_c: $\{T_\eta^{(r)}(\cdot) \mid r = 1, 2, \cdots, R\}$, where each transformation $T_\eta^{(r)}(\cdot)$ with its parameters η corresponds to a specific condition r so that for each condition r the transformed model $T_\eta^{(r)}(\lambda_c)$ gives a better description of the data $X^{(r)}$ collected under this condition r ($r = 1, 2, \cdots, R$). The same algorithm of "speaker-adaptive training" in [1] can be used to estimate the compact model λ_c and the corresponding transformations $T_\eta^{(r)}(\cdot)$ according to the Maximum likelihood criterion. However, in the testing phase, it is not appropriate to use the compact model λ_c to evaluate the testing data directly because λ_c would not match the original data due to the involved transformations. Furthermore, we do not know which transformation should be used for each single testing utterance because we have no idea of which *condition* it comes from. In this paper, the idea of Bayesian Prediction is proposed to solve this problem. In the training stage, a prior distribution of transformation parameters is simultaneously estimated to represent the knowledge of all transformations possibly used in training stage. In the testing phase, the Bayesian Predictive Classification (BPC) algorithm helps to make an optimal decision given the information supplied by the prior distribution on the transformation parameters.

First of all, we must choose a suitable functional form for the transformation $T_\eta^{(r)}(\cdot)$. Obviously, this requires that: i) The transformation is sufficiently powerful to normalize the acoustic difference caused by pronunciation variability; ii) The transformation form is simple enough so that Bayesian prediction is tractable in the decoding phase. One possible choice is the piecewise linear transformation. In this work, as the first step, we choose the simplest transform, namely the bias vector plus the mean vector of HMM: $m'_{ikd} = m_{ikd} + b_d \quad (d = 1, 2, \cdots, D)$.

In principle, each transformation could be related or tied to any different segments of the speech signal. In this work, we assume that each transform is HMM state-dependent; i.e., we use different transformations for different HMM states and the transformations of the state parameters are tied based on the triphone state-tying in the entire HMM set.

Secondly, we need to choose a proper prior distribution for transformation parameters, i.e., b in this case. In order to have a simple form in decoding stage, We choose the following prior p.d.f. based on the concept of natural conjugate prior:

$$\rho(b) = \prod_{d=1}^{D} \sqrt{\frac{\tau_d}{2\pi}} \exp[-\frac{\tau_d}{2}(b_d - \mu_d)^2] \quad (2)$$

where $\{\mu_d, \tau_d \mid d = 1, 2, \cdots, D\}$ are hyperparameters.

As a remark, we also can use a finite mixture form for the prior distribution as in [7] to have a more accurate description of prior information.

3. THE ROBUST TRAINING ALGORITHM

We integrate the above ideas into the conventional acoustic modeling method of large vocabulary speech recognition system, using triphone decision-tree based state tying[9]. Our robust training strategy designed specifically for pronunciation variation is as follows:

1. Build a base-line system based on the conventional HMM (HTK implementation).
2. Transcribe the pronunciation for all speech utterances in the training set (phone recognition with the base-line system) and obtain Viterbi segmentation of each utterance at the HMM's state level; Then label each frame of the MFCC features with its corresponding word w and with the pronunciation id p of the word w in the current utterance.
3. State tying of all triphone models and estimation: build a single decision-tree for each state of phone models based on all data belonging to its corresponding triphones. For each tied state of the tri-phone models (i.e., each leaf node of the each tree):
   (a) Cluster all data in the state into a total of R different conditions according to the different labels of w and p.
   (b) Pool all related data, $X = \{X^{(r)} \mid r = 1, 2, \cdots, R\}$, where $X^{(r)}$ denotes all data under the condition r. Use the state distribution in the current leaf node as the initial estimate of the compact model λ_c for this tied state. Here λ_c is the mixture Gaussian of this tied state, i.e., $\lambda_c = \{w_k, m_k, r_k \mid 1 \le k \le K\}$.
   (c) Given the current λ_c, estimate R transformations $\{T_\eta^{(r)}(\cdot) \mid r = 1, 2, \cdots, R\}$ for each condition r based on $X^{(r)} = \{x_t^{(r)} \mid 1 \le t \le T^{(r)}\}$: for all $d = 1, 2, \cdots, D$ (use $b[d] = 0$ as initialization)

$$b^{(r)}[d] = \frac{\sum_{t=1}^{T^{(r)}} \sum_{k=1}^{K} \xi_t^{(r)}(k) \cdot r_{kd} \cdot (x_{td}^{(r)} - m_{kd})}{\sum_{t=1}^{T^{(r)}} \sum_{k=1}^{K} \xi_t^{(r)}(k) \cdot r_{kd}} \quad (3)$$

   where $\xi_t^{(r)}(k)$ denotes the probability of $x_t^{(r)}$ in mixture component $l_t = k$, i.e., $\xi_t^{(r)}(k) = \Pr(l_t = k \mid x_t^{(r)}, b^{(r)}) = \frac{w_k \cdot \mathcal{N}(x_t^{(r)} \mid m_k + b^{(r)}, r_k)}{\sum_{k=1}^{K} w_k \cdot \mathcal{N}(x_t^{(r)} \mid m_k + b^{(r)}, r_k)}$.
   (d) Re-estimate the compact model λ_c: for $1 \le k \le K$ and $1 \le d \le D$

$$m_{kd} = \frac{\sum_{r=1}^{R} \sum_{t=1}^{T^{(r)}} \xi_t^{(r)}(k) \cdot r_{kd} \cdot (x_{td}^{(r)} - b^{(r)}[d])}{\sum_{r=1}^{R} \sum_{t=1}^{T^{(r)}} \xi_t^{(r)}(k) \cdot r_{kd}} \quad (4)$$

$$r_{kd} = \frac{\sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}} \xi_t^{(r)}(k)}{\sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}} \xi_t^{(r)}(k)\cdot(x_{td}^{(r)} - m_{kd} - b^{(r)}[d])^2} \quad (5)$$

$$w_k = \frac{\sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}} \xi_t^{(r)}(k)}{\sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}}\sum_{k=1}^{K} \xi_t^{(r)}(k)} \quad (6)$$

(e) Goto step (3c) unless some convergence conditions are met.

(f) Estimate the prior distribution of transformation parameters $\eta = \{\mu, \tau\}$ based on the method of moment: for $1 \le d \le D$

$$\begin{aligned}\mu_d &= \frac{\sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}}\sum_{k=1}^{K} \xi_t^{(r)}(k)\cdot b^{(r)}[d]}{\sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}}\sum_{k=1}^{K} \xi_t^{(r)}(k)} \\ &= \frac{\sum_{r=1}^{R} T^{(r)}\cdot b^{(r)}[d]}{\sum_{r=1}^{R} T^{(r)}} \quad (7)\end{aligned}$$

$$\begin{aligned}\tau_d &= \frac{\sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}}\sum_{k=1}^{K} \xi_t^{(r)}(k)}{\sum_{r=1}^{R}\sum_{t=1}^{T^{(r)}}\sum_{k=1}^{K} \xi_t^{(r)}(k)\cdot(b^{(r)}[d] - \mu_d)^2} \\ &= \frac{\sum_{r=1}^{R} T^{(r)}}{\sum_{r=1}^{R} T^{(r)}\cdot(b^{(r)}[d] - \mu_d)^2} \quad (8)\end{aligned}$$

$\{\mu_d, \tau_d\}$ can be tied for all states related to the current leaf node.

4. ROBUST DECODING BASED ON BAYESIAN PREDICTIVE CLASSIFICATION

Based on Bayesian Prediction Classification[6], given observation X, the optimal recognition result $\hat{W}$ is expressed as

$$\begin{aligned}\hat{W} &= \arg\max_W \Pr(W)\cdot\int_\eta \Pr(X \mid T_\eta(\lambda_c), W)\cdot\rho(\eta)\,\mathrm{d}\eta \\ &= \arg\max_W \Pr(W)\cdot\int_\eta \sum_{s,l}\Pr(X \mid s, l, T_\eta(\lambda_c), W)\cdot\rho(\eta)\,\mathrm{d}\eta \\ &\approx \arg\max_W \Pr(W)\cdot\max_{s,l}\int_\eta \Pr(X \mid s, l, T_\eta(\lambda_c), W)\cdot\rho(\eta)\,\mathrm{d}\eta\end{aligned}$$

Given a test utterance $X = (x_1, x_2, \cdots, x_T)$, the compact model λ_c and the prior p.d.f. $\rho(\eta)$ of transformation parameter η, as shown in eq.(2) with hyperparameters estimated from eqs. (7) and (8). The recursive search procedure for *approximately* accomplishing the above equation is described as follows:

(1) Initialization

$$\delta_1(i) = \pi_i\cdot\tilde{b}_i(x_1) \quad 1 \le i \le N \quad (9)$$

$$\psi_1(i) = 0 \quad 1 \le i \le N \quad (10)$$

where

$$\begin{aligned}\tilde{b}_i(x_t) &= \arg\max_{1\le k\le K}\int w_{ik}\cdot\mathcal{N}(x_t|m_{ik}+b_i, r_{ik})\cdot\rho(b_i)\,\mathrm{d}b_i \\ &= \arg\max_{1\le k\le K} w_{ik}\cdot\prod_{d=1}^{D}\sqrt{\frac{\tau_d^{(i)} r_{ikd}}{2\pi(\tau_d^{(i)} + r_{ikd})}} \\ &\exp[-\frac{\tau_d^{(i)} r_{ikd}}{2(\tau_d^{(i)} + r_{ikd})}(x_{td} - m_{ikd} - \mu_d^{(i)})^2]\end{aligned}$$

Here, $\delta_t(i)$ denotes the partial predictive value based on the optimal partial path arriving at state i at the time instant t. The corresponding best partial path is represented by a chain of points started from $\psi_t(i)$.

(2) Recursion: for $2 \le t \le T$, $1 \le j \le N$, do

(2.1) Path-merging in state j:

$$\delta_t(j) = \max_{1\le i\le N}[\,\delta_{t-1}(i)\cdot a_{ij}] \quad (11)$$

$$\psi_t(j) = \arg\max_{1\le i\le N}[\,\delta_{t-1}(i)\cdot a_{ij}] \quad (12)$$

(2.2) Update the partial predictive value:

If (it is the first time to involve state j in computation of $\delta_t(j)$)[1], **then**

$$\delta_t(j) = \delta_t(j)\times\tilde{b}_j(x_t) \quad (13)$$

else

$$\delta_t(j) = \delta_t(j)\times\frac{\tilde{b}_j(x_{j_1}, x_{j_2}, \cdots, x_{j_{L_j}})}{\tilde{b}_j(x_{j_1}, x_{j_2}, \cdots, x_{j_{(L_j-1)}})} \quad (14)$$

where L_j is the accumulated number of feature vectors belonging to state j based on the optimal partial path up to the time instant t; x_{j_i} denotes the ith vector in the state j; and $\tilde{b}_j(x_{j_1}, x_{j_2}, \cdots, x_{j_{L_j}})$ denotes the contribution of data $\{x_{j_1}, x_{j_2}, \cdots, x_{j_{L_j}}\}$, residing in state j, to the partial predictive value $\delta_t(j)$:

$$\begin{aligned}\tilde{b}_j(x_{j_1}, x_{j_2}, \cdots, x_{j_n}) &= \int p(x_{j_1}, x_{j_2}, \cdots, x_{j_n} \mid \\ m_{ik}+b_i, r_{ik})\cdot\rho(b_i)\,\mathrm{d}b_i \quad (15)\end{aligned}$$

(3) Termination

$$\tilde{p}(X|W) \approx \max_i\ \delta_T(i) \quad (16)$$

$$s_T^* = \arg\max_i \delta_T(i) \quad (17)$$

(4) Path (state sequence) Backtracking

$$s_t^* = \psi_{t+1}(s_{t+1}^*) \qquad t = T-1, T-2, \cdots, 1 \quad (18)$$

Here $\tilde{b}_j(x_{j_1}, x_{j_2}, \cdots, x_{j_n})$ is calculated based on the "closest" mixture component label sequence corresponding to the data $\{x_{j_1}, x_{j_2}, \cdots, x_{j_n}\}$:

$$\begin{aligned}\tilde{b}_j(x_{j_1}, x_{j_2}, \cdots, x_{j_n}) &\approx \prod_{k=1}^{K}\omega_{jk}^{L'_k}\cdot\tilde{f}_{jk}(x_{l_1^k}, \cdots, x_{l_{L'_k}^k}) \\ &= \prod_{k=1}^{K}\omega_{jk}^{L'_k}\cdot\prod_{d=1}^{D}\tilde{f}_{jkd}(x_{l_1^k d}, \cdots, x_{l_{L'_k}^k d}) \quad (19)\end{aligned}$$

[1] Including all states tied to state j.

where $\{x_{j_1}, x_{j_2}, \cdots, x_{j_n}\}$ denote feature vectors belonging to state j in X, among which $l_1^k \cdots l_{L'_k}^k$ denote labels of the vectors "closest" to the mixture component k of state j. Then we have

$$\bar{f}_{jkd}(x_{1d}, x_{2d}, \cdots, x_{\zeta d}) = \sqrt{(\frac{r_{jkd}}{2\pi})^{\zeta} \frac{\tau_d^{(j)}}{\zeta r_{jkd} + \tau_d^{(j)}}}$$
$$\exp[-\frac{\zeta r_{jkd}}{2}(\overline{x_\zeta^2} - \overline{x}_\zeta^2)] \exp[-\frac{\zeta r_{jkd}\tau_d^{(j)}}{2(\zeta r_{jkd} + \tau_d^{(j)})}(\overline{x}_\zeta - \mu_d^{(j)})^2]$$

where $\overline{x}_\zeta = \frac{1}{\zeta}\sum_{i=1}^{\zeta}(x_{id} - m_{jkd})$ and $\overline{x_\zeta^2} = \frac{1}{\zeta}\sum_{i=1}^{\zeta}(x_{id} - m_{jkd})^2$.

5. EXPERIMENTS: SWITCHBOARD TASK

In this work, we choose a fast evaluation subset of Switchboard corpora in Workshop 1996 (WS96) at John Hopkins University, which is approximately 10 hours in duration. It is called "10-hr" set hereafter.

5.1. The 10-hr baseline system

In the baseline system, we use the 39-dimension feature vector which is composed of 12 MFCC's with log-energy, and delta and acceleration coefficients. The cepstral normalization is performed in the utterance level. The acoustic models is 3-state 5-mixture-per-state word-internal tri-phone HMM's. The standard phonetic decision tree method is used for state-tying. After tying, the number of all distinct tied-states is reduced to approximately 2K. The dictionary only consists of all words (about 6474 words) occurring in the '10-hr' set. The multiple pronunciation is adopted for some words. The language model is the back-off bigram models which is trained only on the transcriptions of all utterances in the '10-hr' set. The testing data set consists of 200 utterances which are randomly selected from the evaluation test set in WS96.

5.2. Definition and choice of "conditions"

One of the most important implementation issues here is how to define the condition and partition the data into different conditions. It is crucial to have a good tradeoff between the number of conditions and the amount of data used for each condition. In this work, we have tried the two methods to define a *condition*: (I) Align training data in the word level, then perform phone recognition for each word based on the alignment boundary. The phoneme recognition results are viewed as the pronunciation of this word. We usually cluster all different pronunciations of every word into 4 classes or fewer. In training stage (3a), all data from the same word and same pronunciation class are treated as from the same condition. (II) Align training data according to transcriptions,then perform phone recognition for each word based on the alignment boundary. When doing decision-tree state-tying, all data in this state which corresponds to different phoneme recognition results are treated as from different conditions.

5.3. Preliminary results

Some preliminary experimental results are included in Table 5.3. From the results, we can see that the robust training method gives close to 1% reduction in word error rate over a well-trained baseline system. We also see that, for the 10-hr data, the method (II) achieves somewhat better results because method (I) usually causes too many conditions and too little data for each condition.

	Sub	**Del**	**Ins**	**WER**
10-hr baseline	43.94	17.92	3.49	65.39
Robust Training (I)	41.94	18.58	4.31	64.84
Robust Training (II)	42.61	18.17	3.90	64.68

Table 1: The performance (in %) comparison of the new training approach with the 10-hr baseline system

5.4. Discussions

Although we have observed some moderate WER reduction for the Switchboard task, the performance improvement is smaller than what we expected. The possible reasons are: i) Pronunciation variations are very complicated and the associated variability in acoustic realization is only a small part of the problem. ii) The Switchboard task involves many other independent types of variabilities which have not been addressed in this work. iii) The variations also affect recognition in the testing stage, and the robust training strategy has yet to be combined with other methods. In particular, phonetic reduction has been found to be one major cause of variability for spontaneous speech which requires dynamic modeling methodologies beyond the conventional HMMs [3]. Our current robust training strategy will be further developed for new dynamic models of spontaneous speech intended to incorporate phonetic reduction (target undershoot) as well as pronunciation variations.

Acknowledgements: The authors would like to thank Qiang Huo of HKU for useful comments and discussion on this work, and thank Ian Stokes-Rees for help in building the baseline system.

6. REFERENCES

[1] T. Anastasakos, J. McDonough, R. Schwarts and J. Makhoul, "A compact model for speaker-adaptive training," *Proc. ICSLP-96*, pp.1137-1140, 1996.

[2] B. Byrne, M. Finke, S. Khudanpur, J. McDonough, H. Nock, M. Riley, M. Saraclar, C. Wooters, G. Zavaliagkos, "Pronunciation Modeling for conversational speech recognition: A status report from WS97," *Proc. IEEE workshop ASRU-97*, pp.26-33, Nov. 1997.

[3] L. Deng and J. Ma. "A statistical coarticulatory model for the hidden vocal-tract-resonance dynamics," *Proc. Eurospeech*, Budapest, 1999.

[4] Q. Huo and B. Ma, "Irrelevant variability normalization in learning structure from data: a case study on decision-tree based HMM state tying," *Proc. ICASSP-99*, May, 1999.

[5] Q. Huo, Personal communications.

[6] H. Jiang, K. Hirose and Q. Huo, "Robust Speech Recognition based on Bayesian Prediction Approach," *IEEE Trans. on Speech and Audio Processing*, Vol.7, No.4, pp.426–440, July 1999.

[7] H. Jiang, K. Hirose and Q. Huo, "Improving Viterbi Bayesian Predictive Classification via sequential Bayesian leaning in robust speech recognition," *Speech Communication*, Vol. 28, 4, pp.313-326, August 1999.

[8] D. McAllaster, L. Gillick, F. Scattone and M. Newman, "Fabricating conversational speech data with acoustic models: A program to examine model-data mismatch," *Proc. ICSLP-98*, pp.1847-1850, Dec. 1999.

[9] S.J. Young, J.J. Odell and P.C. Woodland, "Tree-based state tying for high accuracy acoustic modeling," *Proc. ARPA Human Language Technology Workshop*, pp.307-312, 1994.