음성인식을 위한 변환 공간 모델에 근거한 순차 적응기법

Sequential Adaptation Algorithm Based on Transformation Space Model for Speech Recognition

김 동 국* · 장 준 혁** · 김 남 수***
Dong Kook Kim · Joo-Hyuk Chang · Nam Soo Kim

ABSTRACT

In this paper, we propose a new approach to sequential linear regression adaptation of continuous density hidden Markov models (CDHMMs) based on transformation space model (TSM). The proposed TSM which characterizes the a priori knowledge of the training speakers associated with maximum likelihood linear regression (MLLR) matrix parameters is effectively described in terms of the latent variable models. The TSM provides various sources of information such as the correlation information, the prior distribution, and the prior knowledge of the regression parameters that are very useful for rapid adaptation. The quasi-Bayes (QB) estimation algorithm is formulated to incrementally update the hyperparameters of the TSM and regression matrices simultaneously. Experimental results showed that the proposed TSM approach is better than that of the conventional quasi-Bayes linear regression (QBLR) algorithm for a small amount of adaptation data.

Keywords : Speech Recognition, Speaker Adaptation, Transformation Space Model, Latent Variable Model, Quasi-Bayes estimate

1. 서 론

현재 자동 음성인식 시스템이 실제 환경의 다양한 응용 분야에서 사용되는 경우 인식 시스템의 성능은 학습과 인식 환경 사이의 불일치(mismatch)로 인하여 현저히 저하된다 [17]. 그러므로 인식 시스템이 성공적으로 사용되기 위해서는 이러한 불일치를 줄이기 위한 다양한 적응 기법이 요구된다. 일반적으로 좋은 적응 알고리즘은 다음과 같이 두 가지 바람직한 특성을 갖는다 [16][18]. 첫째, 적응 알고리즘은 적은 양의 적응 데이터에 대해서도 효과적인 빠른(rapid) 적응 성질을 가져야하며, 데이터 크기가 증가함에 따라 일치된 조건의 성능으로 근사적으로 접근해야 한다. 두 번째는 전에 사용된 적응 데이터를 저장하지 않고 변화하는 환경에 연속적으로 적응할 수 있는 순차 적응 특성

* 전남대학교 전자컴퓨터정보통신 공학부
** 캘리포니아 주립대학, 산타바바라 박사후 연구원
*** 서울대학교 전기컴퓨터 공학부

이다. 순차 적응의 장점은 계산 량과 메모리의 효율성을 가지고 모델을 적응할 수 있다는 것이다.

많은 모델기반 적응 기법들은 새로운 화자의 음향학적 특성들을 더 잘 일치시키기 위해 연속 밀도 은닉 마코프 모델(CDHMM: continuous-density hidden Markov model)의 파라메터를 적응한다 [18][22]. 일반적으로 모델기반 적응 기법은 3 가지로 분류 된다 [22]. 이는 maximum a posteriori (MAP) 기법 [8], maximum likelihood linear regression(MLLR)을 사용한 변환기반 기법 [19] 그리고 화자 공간(speaker space) [7][15] 기법이다. Kuhm [16] 등은 이러한 분류에 따라 빠른 적응 기법들에 대해 고찰하였고, Lee와 Huo [18]는 선 진화(prior evolution) 개념에 기반으로 하는 다양한 순차 적응 기법들을 고찰하였다.

Cluster adaptive training (CAT) [7]와 eigenvoice [15] 같은 화자 공간 적응 기법들은 빠른 화자 적응을 위해 소개되었다. 이러한 기법들은 주요한 화자 모델을 선형 결합에 의해 새로운 화자 모델을 표현함으로서 화자 적응을 수행한다. 주요 화자 모델과 관련된 파라메터는 기본 벡터들의 집합으로 취급되어지며 이를 학습 화자의 선 지식(a priori knowledge)을 나타내는 화자 공간이라 말한다. 최근에 eigenvoice 기법을 Bayesian 적응 구조로 확장하기 위해 factor analysis(FA) [20]와 probabilistic principal component analysis(PPCA) [21]와 같은 은닉 공간 모델(latent variable models)에 근거하여 화자 공간 모델(SSM: speaker space model) [12][13]이라 불리는 빠른 화자 적응 기법이 소개되었다. 은닉 변수 모델은 잘 학습된 화자 종속 모델들로부터 expectation maximization(EM) 알고리즘 [6]을 사용하여 주요 화자 공간 모델을 구한다. 은닉 공간 모델은 CDHMM의 다른 파라메터 사이의 상관관계와 선 확률 분포 정보를 제공해 준다.

MLLR은 maximum likelihood(ML) 기준에 따라 추정된 선형 regression 함수들의 집합을 통해 전체 CDHMM 파라메터들을 적응하는 변환기반 적응 기법이다. MLLR이 빠른 적응에 사용되기 위해 변환 파라메터를 강인하게 추정하는 것이 필요하며, 이를 위해 다양한 기법들이 제안되었다. Maximum a posteriori linear regression (MAPLR) [5]는 변환 파라메터에 대한 선 분포(a priori distribution)를 MAP 추정 기법에 적용함으로서 MLLR 적응 기법을 향상하기 위해 제안되었다. 또한 eigenspace 기반 MLLR과 MAPLR 기법은 PCA [11]와 PPCA을 각각 사용하여 학습 화자와 관련된 변환 행렬들을 분석함으로 제안되었다 [1][2]. SSM과 비슷하게 학습 화자의 선 정보를 특징짓는 변환 공간(transformation space)은 각 화자 종속(SD: speaker-dependent) CDHMM 파라메터 대신에 각 학습 화자에 대한 대표 벡터로 변환 파라메터를 취함으로 정의된다.

많은 순차 적응 기법들은 음성인식 시스템이 새로운 환경에 연속적으로 적응할 수 있도록 연구되어 왔다 [3][9][18]. Quasi-Byes (QB) 추정 [9]에 근거한 다양한 적응 기법들이 순차적으로 CDHMM 파라메터를 갱신하도록 제안되었다. Huo와 Lee [9]은 동시에 CDHMM 파라메터와 hyperparameter을 선 진화 과정을 통해 순차적으로 갱신하는 QB 학습 구조를 적용하였다. 더불어 모든 CDHMM 평균 파라메터가 결합 선 분포 형태의 상관관계를 갖는 경우에 QB 추정 기법을 적용하였다 [10]. Chien [3]은 변환 파라메터에 대한 특별한 선 확률 밀도 함수(pdf: probability density function)를 가정하고 간단한 변환을 수행함으로 순차적인 변환 기반 QB 적응 알고리즘을 제안하였다. 또한 CDHMM의 순차적인 linear regression 적응을 위한 quasi-Byes linear regression (QBLR) 알고리즘을 제안하였다 [4].

이 논문에서는 변환 공간 모델(TSM: transformation space model)에 근거한 순차적인 선형

regression 적응기법을 제안한다. 이러한 기법은 학습 화자의 선 지식을 위해 화자 종속 MLLR 행렬 파라메터를 기반으로 한다. QB 추정 알고리즘은 TSM 의 hyperparameter와 regression 행렬을 동시에 순차적으로 갱신하도록 제안되었다.

이 논문은 다음과 같이 구성된다. 먼저 2 장에서는 FA나 PPCA와 같은 은닉 변수 모델에 기반한 TSM을 소개한다. 3 장에서는 TSM에 근거한 순차 적응 기법을 제안한다. TSM의 QB 학습과 순차적응을 이 장에서 제시된다. TSM을 사용한 순차적인 화자적응에 대한 결과가 4 장에서 제시된다. 마지막으로 5 장에서 결론을 맺는다.

2. 변환 공간 모델

K 개의 mixture 성분을 갖는 N-상태 CDHMM, 즉 $\lambda=\{\lambda\}=\{w_{jk},\ \boldsymbol{\mu}_{jk},\ \boldsymbol{\Sigma}_{jk}\}, j=1,\cdots,N,\ k=1,\cdots,K$을 고려하자. 관측 벡터 $\boldsymbol{x}_t$에 대해 상태종속 확률 밀도 함수는 다중변수 Gaussian 분포의 mixture로 다음과 같이 정의된다.

$$p(\boldsymbol{x}_t|\lambda_j)=\sum_{k=1}^{K} w_{jk} N(\boldsymbol{x}_t|\ \boldsymbol{\mu}_{jk},\ \boldsymbol{\Sigma}_{jk}) \tag{1}$$

여기서 w_{jk}는 상태 j에서 mixture 성분 k에 대한 가중치, $\sum_{k=1}^{K} w_{jk}=1$이며, $\boldsymbol{\mu}_{jk}$는 d 차원 평균 벡터, $\boldsymbol{\Sigma}_{jk}$는 $d\times d$ 공분산 행렬이고 $N(\boldsymbol{x}_t|\ \boldsymbol{\mu}_{jk},\ \boldsymbol{\Sigma}_{jk})$는 다음과 같은 Gaussian 분포이다.

$$N(\boldsymbol{x}_t|\ \boldsymbol{\mu}_{jk},\ \boldsymbol{\Sigma}_{jk}) \propto |\boldsymbol{\Sigma}_{jk}|^{-1/2}\exp\left(-\frac{1}{2}(\boldsymbol{x}_t-\boldsymbol{\mu}_{jk})^T\ \boldsymbol{\Sigma}_{jk}^{-1}(\boldsymbol{x}_t-\boldsymbol{\mu}_{jk})\right) \tag{2}$$

MLLR는 새로운 화자나 환경에 CDHMM 파라메터를 적응하기 위한 가장 효율적인 기법 중에 하나이다. MLLR는 ML 기준에 의해 추정된 선형 regression 함수의 집합에 의해 CDHMM 파라메터가 적응되는 변환 기반 적응 기법이다. Gaussian 분포는 C 개의 regression 클래스로 분류되고 regression 클래스 c와 관련된 변환 행렬 $\boldsymbol{W}^{(c)}$는 같은 클래스에 속하는 모든 분포에 의해 공유된다. MLLR 적응 기법은 $d\times(d+1)$ 차원 클러스터 종속적인 regression 행렬 $\boldsymbol{W}^{(c)}$을 $(d+1)\times 1$ 차원 확장된 평균 벡터 $\zeta_{jk}=[1,\ \boldsymbol{\mu}_{jk}{}^T]^T$에 적용하여 CDHMM 평균 파라메터 $\boldsymbol{\mu}_{jk}$를 다음과 같이 적응한다.

$$\hat{\boldsymbol{\mu}}_{jk}=\boldsymbol{W}^{(c)}\zeta_{jk} \tag{3}$$

이 때 적응된 평균 벡터에 대한 $\boldsymbol{x}_t$의 관측 유사도는 다음과 같다.

$$N(\boldsymbol{x}_t|\ \boldsymbol{W}^{(c)},\ \boldsymbol{\mu}_{jk},\ \boldsymbol{\Sigma}_{jk}) \propto |\boldsymbol{\Sigma}_{jk}|^{-1/2}\exp\left(-\frac{1}{2}(\boldsymbol{x}_t-\boldsymbol{W}^{(c)}\zeta_{jk})^T\boldsymbol{\Sigma}_{jk}^{-1}(\boldsymbol{x}_t-\boldsymbol{W}^{(c)}\zeta_{jk})\right) \quad (4)$$

단순화를 위해 하나의 전체 regression 행렬 $\boldsymbol{W}$을 모든 CDHMM 파라메터에 사용된다고 가정한다.

$\{\boldsymbol{W}_r\}, r\in\{1,\cdots,R\}$는 MLLR 기법을 사용해 얻어진 R 개의 잘 학습된 화자 종속 regression 행렬의 집합이라 하자. 또한 $\boldsymbol{w}_r=[W_{r1},\cdots,W_{rd}]$을 regression 행렬 $\boldsymbol{W}_r$의 행 벡터 $\{W_{ri}\}$을 연결함으로 만들어진 $D(=d(d+1)\times1)$ 차원의 변환 supervector라 하자. 화자 종속 변환 파라메터의 집합 $\boldsymbol{w}=[\boldsymbol{w}_r,\cdots,\boldsymbol{w}_R]$가 FA와 같이 파라메터를 $\theta=\{\boldsymbol{V},\ \overline{\boldsymbol{w}},\ \boldsymbol{\Psi}\}$ 갖는 은닉 변수 모델에 의해 다음과 같이 발생된다고 가정한다.

$$\boldsymbol{w}=\boldsymbol{V}\boldsymbol{z}+\overline{\boldsymbol{w}}+\varepsilon \quad (5)$$

여기서 $\overline{\boldsymbol{w}}$은 변환 supervector의 평균, $\boldsymbol{V}=[\boldsymbol{v}_1,\cdots,\boldsymbol{v}_P]$는 변환 파라메터의 준공간(subspace)을 나타내는 $D\times P$ 차원 행렬, $\boldsymbol{z}=[z_1,\cdots,z_P]$는 $p(\boldsymbol{z})\sim N(0,\boldsymbol{I})$인 P차원 은닉 변수 그리고 ε는 대각 공분산 행렬 $\boldsymbol{\Psi}$을 갖고 $p(\varepsilon)\sim N(0,\boldsymbol{\Psi})$인 $\boldsymbol{z}$에 독립적인 Gaussian random 변수이다. PPCA는 잡음 공분산 행렬이 isotripic, 즉 $\boldsymbol{\Psi}=\delta^2\boldsymbol{I}$로 정의된다. 이러한 가정에 근거하여 $\boldsymbol{z}$가 주어진 경우 $\boldsymbol{w}$에 대한 조건부 행렬은 $p(\boldsymbol{w}|\boldsymbol{z})\sim N(\boldsymbol{V}\boldsymbol{z}+\overline{\boldsymbol{w}},\ \boldsymbol{\Psi})$로 유도될 수 있다. 또한 $\boldsymbol{w}$에 대한 선 분포는 $p(\boldsymbol{w})\sim N(\overline{\boldsymbol{w}},\ \boldsymbol{V}\boldsymbol{V}^T+\boldsymbol{\Psi})$로 구할 수 있다. 은닉 변수 모델의 파라메터 θ는 ML 기준에 의해 추정될 수 있다. 그러나 은닉 변수 $\boldsymbol{z}$는 은닉되어 있기 때문에 θ를 추정하기 위해서는 반복적으로 파라메터를 갱신하는 EM 알고리즘 [6]을 적용한다.

3. TSM에 근거한 순차 적응

3.1 TSM을 위한 QB 학습

이 장에서는 TSM을 위한 QB 학습의 기본 개념과 공식에 대해 고찰한다 [4][9]. $\boldsymbol{X}^n=\{\boldsymbol{X}_1,\ \boldsymbol{X}_2,\cdots,\ \boldsymbol{X}_n\}$은 regression 행렬 $\boldsymbol{w}$와 CDHMM 파라메터 λ와 관련된 iid(independent identically distributed) 관측 벡터 열이라 하자. $\boldsymbol{w}$의 posteriori pdf에 대한 반복적인 표현은 다음과 같다.

$$p(\boldsymbol{w}|\ \boldsymbol{X}^n,\lambda)=\frac{p(\boldsymbol{X}^n|\ \boldsymbol{w},\lambda)\cdot p(\boldsymbol{w}|\ \boldsymbol{X}^{n-1},\lambda)}{\int p(\boldsymbol{X}^n|\ \boldsymbol{w},\lambda)\cdot p(\boldsymbol{w}|\ \boldsymbol{X}^{n-1},\lambda)d\boldsymbol{w}} \quad (6)$$

이 식은 $\boldsymbol{w}$에 대한 반복적인 Bayesian 추정을 하도록 하는 기본적인 식이다. 그러나 반복적인 Bayesian 추정 기법의 이러한 형태의 구현은 매우 어렵기 때문에 이를 해결하기 위해 QB 학습이

라 불리는 기법이 제안되었다 [9]. QB 학습 과정은 각 단계에서 posterior density $p(\boldsymbol{w}|\boldsymbol{X}^{n-1},\lambda)$가 가장 가까운 파라메터 형태의 선 밀도 $g(\boldsymbol{w}|\theta^{(n-1)})$에 의해 두 밀도가 같은 모드(mode)를 갖도록 근사화 된다. $\theta^{(n-1)}$는 앞 데이터 $\boldsymbol{X}^{n-1}$로부터 유도된 진화된 파라메터이다. 시간 n에서 관측 벡터 집합 $\boldsymbol{X}_n$와 근사적인 선 pdf $g(\boldsymbol{w}|\theta^{(n-1)})$가 주어졌다 가정하자. 변환 파라메터 $\boldsymbol{w}$가 hyperparameter $\theta^{(n-1)}$을 갖는 은닉 변수 $\boldsymbol{z}$에 의해 식 (5)에 의해 주어진 모델에 의해 발생됐다고 가정했기 때문에 $\boldsymbol{w}$에 대한 QB 추정은 쉽게 정의된다. ($\boldsymbol{X}_n$, $\boldsymbol{S}_n$, $\boldsymbol{L}_n$)을 $\boldsymbol{X}_n$에 대한 completer-data라 하자. 여기서 $\boldsymbol{S}_n=\{s_t^{(n)}\}$는 상태 열을 나타내며, $\boldsymbol{L}_n=\{l_t^{(n)}\}$은 mixture 성분 열을 나타낸다. 여기서 현재 추정 $\boldsymbol{w}^{(n)}$의 근사적인 posterior density을 갱신할 수 있으며 다음과 같은 EM 단계를 반복함으로서 새로운 추정치 $\hat{\boldsymbol{w}}$을 유도할 수 있다.

E-단계: 다음 보조 함수를 계산하라.

$$R_{QB}(\boldsymbol{w}|\boldsymbol{w}^{(n)},\theta^{(n-1)})=E\left[\log p(\boldsymbol{X}_n,\boldsymbol{S}_n,\boldsymbol{L}_n|\boldsymbol{w},\lambda)+\rho\log g(\boldsymbol{w},\boldsymbol{z}|\theta^{(n-1)})|\boldsymbol{X}_n,\boldsymbol{w}^{(n)}\right] \tag{7}$$

여기서 $0<\rho\leq 1$는 새로운 데이터 $\boldsymbol{X}^n$에 대해 과거 관측 열 $\boldsymbol{X}^{n-1}$의 효과를 줄이기 위한 forgetting factor이다.

M-단계: 파라메터 $\boldsymbol{w}$을 다음에 따라 갱신한다.

$$\hat{\boldsymbol{w}}=\arg\max_{\boldsymbol{w}} R_{QB}(\boldsymbol{w}|\boldsymbol{w}^{(n)},\theta^{(n-1)}) \tag{8}$$

식 (7)과 (8)의 EM-단계를 반복적으로 적용함으로서 근사적인 posterior density는 감소하지 않는다는 것이 증명되었다 [6].

3.2 TSM 기반 순차 적응 유도

선 분포의 선택은 QB 추정을 위한 가장 중요한 문제 중에 하나이다. Regression 행렬의 선 밀도는 posterior distribution와 같은 모드를 갖도록 conjugate distribution family에 속하도록 선택되어야 한다. 이 논문에서는 식 (5)의 TSM에 근거하여 regression 행렬의 선 밀도는 다음과 같이 normal density로 정의된다.

$$\begin{aligned} g(\boldsymbol{w}^{(n)}|\theta^{(n-1)}) \propto & |\boldsymbol{V}^{(n-1)}\boldsymbol{V}^{(n-1)T}+\boldsymbol{\Psi}^{(n-1)}|^{-1/2}\cdot \\ & \exp\left(-\frac{1}{2}(\boldsymbol{w}^{(n)}-\overline{\boldsymbol{w}}^{(n)})^T(\boldsymbol{V}^{(n-1)}\boldsymbol{V}^{(n-1)T}+\boldsymbol{\Psi}^{(n-1)})^{-1}(\boldsymbol{w}^{(n)}-\overline{\boldsymbol{w}}^{(n)})\right) \end{aligned} \tag{9}$$

위의 TSM과 관련된 hyperparameters $\theta^{(n-1)}=\{\boldsymbol{V}^{(n-1)},\overline{\boldsymbol{w}}^{(n-1)},\boldsymbol{\Psi}^{(n-1)}\}$ 는 앞에서 등록된 데이터로부터 얻을 수 있다. 그러면 E-단계 (7)에서 정의된 보조 함수는 다음과 같이 다시 쓰여진다.

$$R_{QB}(\boldsymbol{w}|\boldsymbol{w}^{(n)},\theta^{(n-1)}) = \sum_{S_n}\sum_{L_n} p(S_n, L_n|X_n, \boldsymbol{w}^{(n)},\lambda)\log p(X_n, S_n, L_n \boldsymbol{w},\lambda) + \rho\log g(\boldsymbol{w}, z|\theta^{(n-1)})|X_n, \boldsymbol{w}^{(n)}] \quad (10)$$

조건부 확률 분포를 대입하므로 보조 함수는 다음과 같이 다시 기술되어진다.

$$R_{QB}(\boldsymbol{w}|\boldsymbol{w}^{(n)},\theta^{(n-1)}) \propto -\frac{1}{2}\sum_t\sum_{j,k}\gamma_t(j,k)\left((x_{t^{(n)}}-W\xi_{jk})^T\Sigma_{jk}^{-1}(x_{t^{(n)}}-W\xi_{jk})\right) + \rho E\left(-\frac{1}{2}(\boldsymbol{w}-V^{(n-1)}z-\overline{w}^{(n-1)})^T\Psi^{-1,(n-1)}(\boldsymbol{w}-V^{(n-1)}z-\overline{w}^{(n-1)})|\boldsymbol{w}^{(n)}\right) \quad (11)$$

여기서 $\gamma_t(j,k)=P(s_t^{(n)}=j, l_t^{(n)}=k|X_n, \boldsymbol{w}^{(n)})$이다. 여기서 다음과 같은 확장된 평균 행렬 C_{jk}을 정의한다.

$$C_{jk}=\begin{bmatrix} 1 & \mu_{jk1} & \cdots & \mu_{jkd} & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & \mu_{jk1} & \cdots & \mu_{jkd} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \cdots & 1 & \mu_{jk1} & \cdots & \mu_{jkd} \end{bmatrix} \quad (12)$$

이를 이용하여 보조 함수는 다음과 같이 다시 쓰여진다.

$$R_{QB}(\boldsymbol{w}|\boldsymbol{w}^{(n)},\theta^{(n-1)}) \propto -\frac{1}{2}\sum_t\sum_{j,k}\gamma_t(j,k)\left((x_{t^{(n)}}-C_{jk}w)^T\Sigma_{jk}^{-1}(x_{t^{(n)}}-C_{jk}w)\right) + \rho E\left(-\frac{1}{2}(\boldsymbol{w}-V^{(n-1)}z-\overline{w}^{(n-1)})^T\Psi^{-1,(n-1)}(\boldsymbol{w}-V^{(n-1)}z-\overline{w}^{(n-1)})|\boldsymbol{w}^{(n)}\right) \quad (13)$$

위 수식을 전개한 후 위 식은 정규화 상수 C만큼 곱해진 지수 함수는 다음과 같은 normal 분포 형태로 표현된다.

$$C\cdot\exp\left(R_{QB}(\boldsymbol{w}|\boldsymbol{w}^{(n)},\theta^{(n-1)})\right) \propto \exp\left(-\frac{1}{2}(\boldsymbol{w}-\hat{m})^T\hat{\Theta}^{-1}(\boldsymbol{w}-\hat{m})\right) \quad (14)$$

여기서 새로운 hyperparameter $\hat{m}$과 $\hat{\Theta}$는 다음과 같이 정의된다.

$$\hat{m}=\left(\sum_t\sum_{j,k}\gamma_t(j,k)C_{jk}^T\Sigma_{jk}^{-1}C_{jk}+\rho\,\Psi^{-1,(n-1)}\right)^{-1}\cdot \left(\sum_t\sum_{j,k}\gamma_t(j,k)C_{jk}^T\Sigma_{jk}^{-1}C_{jk}\,x_t^{(n)}+\rho\,\Psi^{-1,(n-1)}(V^{(n-1)}\hat{z}+\overline{w}^{(n-1)})\right)^{-1} \quad (15)$$

$$\hat{\Theta}=\left(\sum_t\sum_{j,k}\gamma_t(j,k)C_{jk}^T\Sigma_{jk}^{-1}C_{jk}+\rho\,\Psi^{-1,(n-1)}\right)^{-1} \quad (16)$$

그리고 여기서 $\hat{z}$는 다음과 같다.

$$\begin{aligned}\hat{z} &= E[z|\, w^{(n)}] \\ &= \left(I + V^{T,(n-1)}\, \Psi^{-1,(n-1)}\, V^{(n-1)}\right) V^{T,(n-1)}\, \Psi^{-1,(n-1)} \left(w^{(n)} - \overline{w}^{(n-1)}\right)\end{aligned} \tag{17}$$

Complete data의 근사적인 posterior density는 갱신된 hyperparameter $\hat{m}$와 $\hat{\theta}$을 갖고 식 (9) $g(w^{(n)}|\theta^{(n-1)})$와 같은 normal distribution family에 속한다. 변환 공간는 regression 행렬에 대한 학습 화자의 선 지식과 관련되기 때문에 적응 데이터를 관찰한 후에도 변화되지 않는다. 그러므로 hyperparameter V는 선 진화 동안에 고정되었다고 가정하고 Ψ에 대한 선 진화는 $\hat{\theta}$에 의해 근사화 된다. 결과적으로 TSM에 의한 선 진화는 다음과 같이 기술된다.

$$\overline{w}^{(n)} = \hat{m} \tag{18}$$

$$\Psi^{(n)} = \hat{\theta} \tag{19}$$

$$V^{(n)} = V^{(n-1)} \tag{20}$$

PPCA의 경우는 비슷한 결과, 즉 $\overline{w}^{(n)} = \hat{m}$, $\Psi^{(n)} = \delta^{2,(n-1)} I$을 갖고 $\delta^{2,(n)} = trace\{\hat{\theta}\}/D$으로 얻어진다. TSM 진화 과정이 마친 후에 시간 n에서 QB 추정된 변환 파라메터 $\hat{w}^{(n)}$은 다음과 같이 진화된 선 pdf의 모드를 취함으로 얻어진다.

$$\hat{w}^{(n)} = \hat{m}\,. \tag{21}$$

3.3 다른 적응 기법과 비교

제안된 TSM 진화 기법은 QBLR 기법과 비슷하다 [4]. QBLR 구조 하에서는 $d\times(d+1)$ 차원 regression 행렬 W는 MAPLR에서 제안된 matrix variate normal distribution에 의해 모델링되었다. Matrix variate normal distribution는 D 차원을 갖는 평균 벡터 $\hat{m}$와 대각 block 요소가 $(d+1)\times(d+1)$ 차원 행렬을 가진 $d(d+1)\times d(d+1)$ 차원의 block 대각 matrix $\Delta^{(n-1)}$ 갖는 univariate 형태 $p(w|\theta^{(n-1)}) \sim N(\overline{w}^{(n-1)}, \Delta^{(n-1)})$로 표현될 수 있다. 이러한 분포는 hyperparameter로서 평균과 공분산 행렬, 즉 $\theta^{(n-1)} = (\overline{w}^{(n-1)}, \Delta^{(n-1)})$을 갖는다. 위의 선 진화에 근거하여 QBLR 알고리즘은 regression 행렬 W가 추정되고 선 통계치 $\theta^{(n)}$가 QB 학습에 의해 갱신하도록 하는 순차 적응 기법이 개발되었다. [4]

만약 식 (15)에서 $\hat{z}=0$라 놓으면, TSM 진화는 QBLR과 같은 식이 된다. QBLR 기법과 비교하여 식 (15)에 있는 TSM 진화 기법은 갱신된 평균 hyperparameter $\hat{m}$가 변환 공간 안에서 추정된 선 변환 모델이 QBLR 추정 과정에 통합된 형태로 나타남을 알수 있다.

Eigenspace 기반 MAPLR 기법은 PPCA에 근거하여 regression 행렬 w을 MAP 추정하기 위해 일괄 형태의 적응 기법을 수행하도록 제안되었다. Eigenspace 기반 MAPLR 기법은 $\Psi = \delta^2 I$을 갖

고 식 (10)과 같은 방식으로 정의된다. 그러므로 regression 파라메터에 대한 eigenspace 기반 MAPLR 해는 forgetting factor을 1을 갖는 식 (21)과 같다. 이것은 TSM 진화 기법이 eigenspace 기반 MAPLR을 특별한 경우로 하는 일반적인 적응기법임을 나타낸다.

4. 실험 및 결과고찰

4.1 데이터베이스와 인식 시스템

제안된 화자 적응 알고리즘을 성능을 평가하기 위해 과학원에서 제공된 두 개의 다른 음성 데이터베이스를 사용하였다. 첫 번째는 한국어 연결 숫자음 데이터베이스이고 다른 것은 대용량 어휘 연속 음성인식(LVCSR: large vocabulary continuous speech recognition)인 무역 상담용 3000 단어 음성 데이터베이스이다.

첫 번째 적응 task을 위해 105 명 화자 (남자 68 명, 여자 37 명)로부터 19891 숫자로 구성된 3880 발음이 학습을 위해 사용되었고 35 명 화자 (남자 25 명, 여자 13 명)로부터 4815 숫자로 구성된 939 발음이 평가를 위해 사용되었다. 각각의 화자는 3-7 개의 숫자로 구성된 30-40 개의 문장을 발음하였고 각 문장은 평균 1.3 초로 구성되었다. 각 숫자는 7 개 상태를 갖는 skip이 없는 left-to-right HMM에 의해 모델링 되었다. 4 개의 mixture Gaussian은 각 상태의 관측 분포를 표현하기 위해 사용되었다. 인식 실험에서는 적응을 위해 각 목적 화자로부터 10 개의 문장을 사용하였고 나머지 문장에 대해 인식 실험을 수행하였다. 음성 신호는 8 kHz로 샘플링되었고 20 ms overlap을 갖고 매 10 ms마다 30 ms 프레임으로 분할되었다. 각 프레임은 12 차 mel-frequency cepstral coefficients와 1 차 도함수로 구성된 총 24 차 특징벡터에 의해 구성되었다. 4 개의 mixture Gaussian을 갖는 화자 독립 시스템은 91.7%의 단어 인식률을 나타내었다.

두 번째 적응 task을 위해 기본 인식기는 120 명 화자 (남자 80 명, 여자 40 명)에 의해 발음된 11,760 문장 (100,408 단어)로 구성된 학습 데이터로 학습 하였고, 10 명 화자 (남자 6 명, 여자 4 명)에 의해 발음된 895 문장 (7,591 단어)로 구성된 문장이 평가를 위해 사용되었다. 각 테스트 화자는 98 문장을 발음하였고, 그 중 10 개는 적응 데이터를 위해 사용되고 나머지 88 문장은 테스트 데이터로 사용되었다. 각 적응 문장은 평균 8.4 단어로 구성되었고 평균 3 초 길이로 구성되었다. Context-dependent 음향 모델이 결정 트리 state tying 알고리즘에 근거하여 구성되었고 약 전체 2,000 개의 다른 상태 클러스터를 형성하였다. 4 개의 Gaussian이 각 상태에서 관측 분포를 표현하기 위해 사용되었다. 단어 클래스에 근거한 bigram 모델이 언어 모델을 위해 사용되었다. 4 개의 mixture Gaussian을 갖는 기본 시스템은 83.8%의 단어 인식률을 나타내었다.

4.2 TSM의 학습

FA와 PPCA에 의해 TSM을 학습하기 위해 먼저 모든 학습 화자로부터 음성을 가지고 화자 독립 모델을 학습하였다. 그리고 SD regression 행렬들은 block-diagonal 행렬을 갖는 일반적인 MLLR 적응 기법을 각각의 학습 화자에 적용하여 얻었다. 유용한 적응 데이터의 양을 고려하여 단지 하나의 전체 regression 클래스만이 MLLR 파라메터 추정을 위해 사용되었다. 묵음 모델은 적응

되지 않았고 묵음 데이터는 적응 과정동안에 사용되지 않았다. HMM 파라메터의 12 차 cepstrum 과 12차 delta- cepstrum 모두 MLLR에 의해 적응되었다. 결과적으로 두 데이터베이스에서 $D=\{13\times12\times2\}=312$ 차원의 변환 supervector가 사용되었다. FA와 PPCA을 위한 파라메터 추정을 위해 EM 알고리즘을 이용하여 차원 $P=10\sim50$을 갖고 $\{\boldsymbol{V}, \overline{\boldsymbol{w}}, \boldsymbol{\Psi}\}$을 구하였다. EM 알고리즘의 수렴 후에 초기 hyperparameter $\theta^{(0)}=\{\boldsymbol{V}^{(0)}, \overline{\boldsymbol{w}}^{(0)}, \boldsymbol{\Psi}^{(0)}\}$은 EM 알고리즘에 의해 주어진 값을 사용하였다.

4.2 기존 알고리즘과 비교

먼저 평균 파라메터만이 적응되는 경우에 여러 가지 일괄 및 순차 적응 기법들에 대한 인식 성능을 비교하였다. 5 가지 다른 기법, 즉 1) MAPLR, 2) eigenvoice, 3) QBLR, 4) PPCA에 의한 TSM 5) FA에 의한 TSM을 사용하여 교사 적응 실험을 수행하였다.

MAPLR 선 분포를 구하기 위해 hyperparameter $\overline{\boldsymbol{w}}$와 $\boldsymbol{\Delta}$은 MLLR 적응 과정에서 얻어진 SD regression 행렬의 ensemble 평균과 공분산을 취함으로서 각각 얻어졌다. Eigenvoice 기법을 위한 화자 basis를 얻기 위해 평균 파라메터에 대해 MLLR과 MAP이 결합된 적응 기법을 사용하여 SD HMM을 학습하였다. 이렇게 학습된 SD 모델을 HMM 평균 벡터들을 모아 supvervector를 구성하였다. 그리하여 supervector는 숫자음에 대해 $D=7680$차원, 대용량 task에 대해서는 $D=191136$차원을 구성하였다. Eigenvoice을 위한 화자 basis는 $P=20$을 갖고 모든 supervector에 PCA을 적용함으로 얻어졌다.

MAPLR 기법을 위해 얻어진 hyperparameter는 QBLR 초기 hyperparameter을 위해 적용되었다. MAPLR에 의해 주어진 변환 파라메터의 추정치는 TSM 진화의 초기 값으로 사용되었다. 순차 적응을 위해 파라메터는 각 적응 문장에 대해 갱신되었다. 순차 적응의 두 task에서 forgetting factor $\rho=1$을 갖고 계산되었다. Forgetting factor가 인식 성능에 큰 영향을 미치지 못하므로 여러 가지 다른 값에 대해서는 나타내지 않았다.

그림 1과 2는 숫자음 인식과 LVCSR task에 대한 MAPLR, eigenvoice, QBLR 그리고 TSM evolution에 대한 성능을 각각 나타낸다. 단어 인식률은 적응 문장의 수에 대해 나타냈다. 여기서 "EV (P=20)"은 화자 basis가 $P=20$을 갖고 얻어진 eigenvoice 기법을 표현한다. 한편, TSM evolution"(PPCA, $P=30$)"과 "TSM evolution (FA, $P=50$)은 각각 $P=30$와 $P=50$을 갖고 PPCA와 FA 진화 기법을 나타낸다.

그림으로부터 MAPLR 기법은 적은 량의 적응 데이터에 대해 좋은 성능 향상을 제공해주지 못하지만 많은 적응 데이터에 따라 더 좋은 결과가 나타났다. Eigenvoice 기법은 매우 적은 량의 적응 데이터에 대해 매우 좋은 결과가 나타났으나 더 많은 데이터가 유용한 경우에 성능이 향상되지 못하였다. QBLR 기법은 숫자음 인식 task에 대해 일괄 MAPLR 기법과 거의 같은 적응 성능을 보였다. LVCSR task에 대해 MAPLR은 더 많은 적응 데이터에 대해 QBLR에 비해 더 좋은 결과를 나타내었다. 한 문장 적응 데이터의 경우 QBLR의 결과는 일괄 MAPLR과 같은 결과를 나타내었다.

MAPLR 기법과 비교하여 TSM 진화 기법은 특별히 한 문장 적응 데이터의 경우에 향상된 인식 성능을 나타내었다. 숫자음 인식 task에서 eigenvoice의 경우와 비교하면 TSM 진화는 두 문장

까지 비슷한 인식 성능을 나타냈었으나 더 많은 적응 데이터의 경우에 더 나은 결과를 나타내었다. LVCSR 경우에는 eigenvoice 기법이 매우 적은 량의 적응 데이터의 경우에 TSM 진화에 비해 약간 더 나은 인식 성능을 나타내었다. 그러나 적응 데이터가 증가함에 따라 TSM 진화가 더 나은 성능을 나타내었다. 이것은 TSM 진화 적응 기법이 매우 적은 크기뿐 아니라 LVCSR task에 대해 매우 효과적임을 나타낸다.

QBLR과 TSM 진화에 근거한 순차 적응 기법들의 인식 성능은 또한 그림 1과 2에 비교되었다. 그림으로부터 TSM 진화 기법은 두 task에 대해 모두 QBLR을 사용하는 경우보다 더 나은 성능을 보였다. 이러한 인식률의 향상은 빠른 적응을 위해 유용한 다양한 정보원을 포함하는 TSM을 사용하기 때문이다. 결과로부터 TSM 진화 기법이 QBLR 기법보다 우수함을 나타내었다.

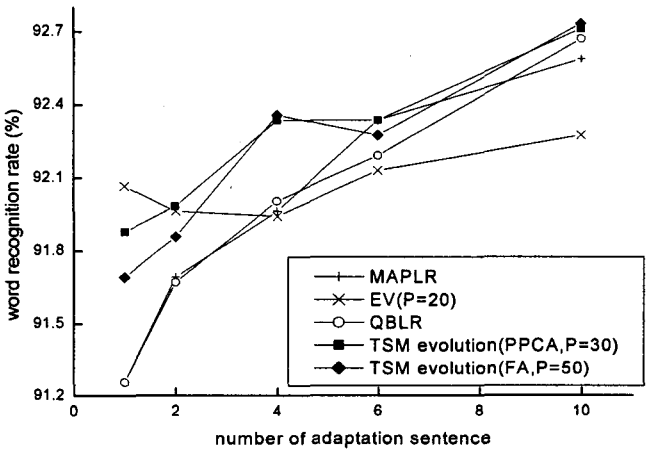

그림 1. 숫자음 인식 시스템에서 MAPLR, eigenvoice, QBLR, TSM 진화 적응 기법에 대한 단어 인식률 비교 (화자독립: 91.7%)

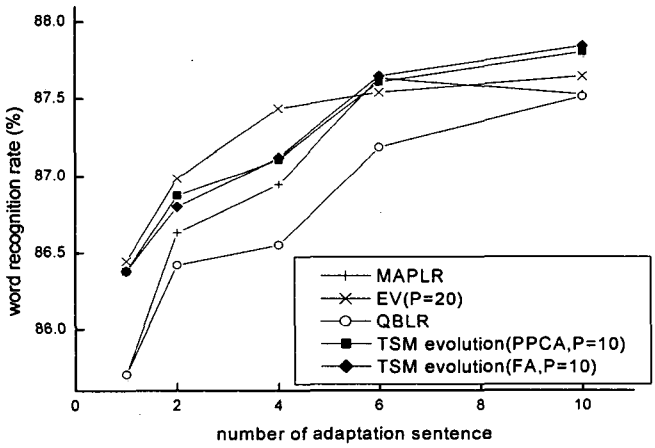

그림 2. LVCSR 시스템에서 MAPLR, eigenvoice, QBLR, TSM 진화 적응기법에 대한 단어 인식률 비교 (화자독립: 83.8%)

4.2 Batch TSM과 비교

그림 3과 4는 다양한 적응 데이터의 량의 따라서 TSM 일괄 및 순차 적응 기법의 성능을 나타낸다. 그림 3으로부터 PPCA 순차 적응 기법은 일괄 적응의 결과와 거의 비슷한 적응 성능을 나타내었고 반면 FA는 일괄 적응이 더 많은 적응 데이터가 주어짐에 따라 더 나은 성능을 나타내었다. LVCSR 경우 제안된 기법의 순차 적응 기법은 모든 적응 조건하에서 일괄 기법과 거의 같은 적응 성능을 나타내었다. 실험 결과로부터 PPCA와 FA에 근거한 TSM 일괄 및 순차 적응 기법은 거의 똑같은 성능을 나타내었다. 또한 일괄 TSM 적응을 그림 1과 2에 있는 MAPLR과 비교하는 경우 TSM 기법이 모든 적응 문장에 대해 MAPLR 기법보다 더 나은 인식 성능을 나타내었다.

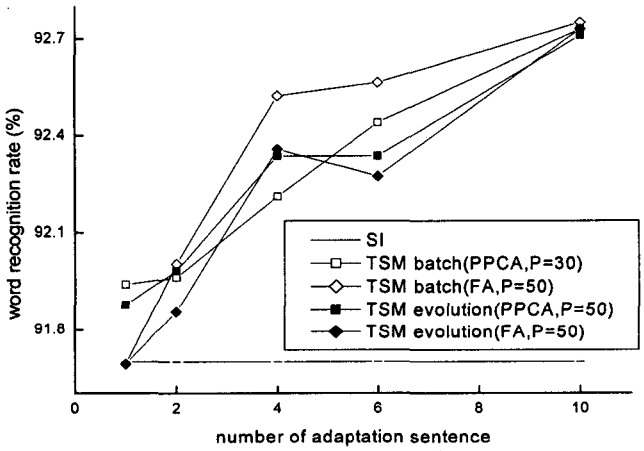

그림 3. 숫자음 음성인식 시스템에서 TSM 기반 기법의 교사 일괄 및 순차 적응의 성능

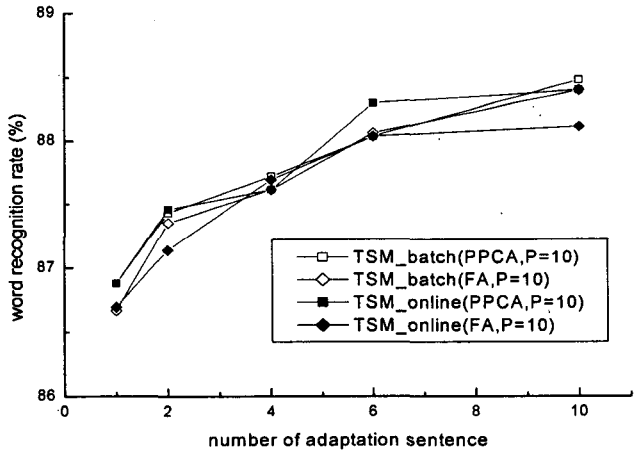

그림 4. LVCSR 시스템에서 TSM 기반 기법의 교사 일괄 및 순차 적응의 성능

5. 결 론

본 논문에서는 CDHMM의 순차적인 선형 regression 적응을 위한 TSM 진화 알고리즘을 제시하였다. FA와 PPCA에 근거한 TSM은 regression 파라메터와 관련된 상관관계 정보, 선 분포 그리고 학습 화자의 선 지식 등과 같은 빠른 적응을 위해 유용한 다양한 정보원을 제공한다. 선 분포의 hyperparameter을 갱신하는 QBLR 기법과 비교하여 TSM 진화는 QB 학습에 근거하여 TSM의 hyperparameter을 순차적으로 갱신하도록 제안되었다. 제안된 알고리즘의 효과성은 한국어 연결 숫자음 인식과 LVCSR task에 대해 실험되었다. 실험 결과 일과 TSM 적응 기법은 두 인식 task에 대해 MAPLR 기법보다 더 나은 성능을 나타내었다. 순차 적응 기법인 TSM 진화는 다양한 적응 데이터에 대해 숫자음 뿐 아니라 LVCSR task에 대해 QBLR보다 더 나은 인식률을 나타내었다.

참 고 문 헌

[1] Chen, K.-T., Liau, W.-W., Wang, H.-M. and Lee, L.-S. 2000. "Fast speaker adaptation using eigenspace-based maximum likelihood linear regression." in Proc. Int. Conf. Spoken Language Processing, Beijing, China, pp. 742-745, Oct.

[2] Chen, K.-T. and Wang, H.-M. 2001. "Eigenspace-based maximum a posteriori linear regression for rapid speaker adaptation." in Proc. IEEE Int. Conf. Acoustics, Speech, Singnal Processing, Salt Lake City, USA, vol. 1, pp. 317-320, May.

[3] Chien, J.-T. 1999. "On-line hierarchical transformation of hidden Markov models for speech recognition." IEEE Trans. Speech and Audio Proc., vol. 7, pp. 656-667, Nov.

[4] Chien, J.-T. 2002. "Quasi-Bayes linear regression for sequential learning of hidden Markov models." IEEE Trans. Speech and Audio Proc., vol. 10, pp. 268-278, July.

[5] Chou, W. 1999. "Maximum a posteriori linear regression with elliptically symmetric matrix priors." in Proc. Euro. Conf. Speech Commun., Technology, vol. 1, pp. 1-4.

[6] Dempster, A. P., Laird, N. M. and Rubin, D. B. 1977. "Maximum likelihood from incomplete data via the EM algorithm." Journal of the Royal Statistical Society, vol. 39, pp. 1-38.

[7] Gales, M. J. F. 2000. "Cluster adaptive training of hidden Markov models." IEEE Trans. Speech and Audio Proc., vol. 8, no. 4, pp. 417-428.

[8] Gauvain, J. L. and Lee, C.-H. 1994. "Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains." IEEE Trans. Speech and Audio Proc., vol. 2, pp. 291-298.

[9] Huo, Q. and Lee, C.-H. 1997. "On-line adaptive learning of the continuous density hidden Markov model based on approximate recursive Bayes estimate." IEEE Trans. Speech and Audio Proc., vol. 5, pp. 161-172, Mar.

[10] Huo, Q. and Lee, C.-H. 1998. "On-line adaptive learning of the correlated continuous density hidden Markov models for speech recognition." IEEE Trans. Speech and Audio Proc.), vol. 6, pp. 386-397, July.

[11] Jolliffe, I. T. 1986. Principal component analysis, Springer-Verla.

[12] Kim, D. K. and Kim, N. S. 2001. "Rapid speaker adaptation using probabilistic principal component analysis." IEEE Signal Processing Letters, vol. 8, no. 6, pp. 180-183, June.

[13] Kim, D. K. and Kim, N. S. 2002. "Online adaptation of continuous density hidden Markov models based on speaker space model evolution." in Proc. Int. Conf. Spoken Language Processing, Denver, USA, pp. 1393-1396.
[14] Kim, D. K, Kim, Y. J., Lim, W. H. and Kim, N. S. 2003. "Online adaptation using transformation space model evolution." in Proc. IEEE Int. Conf. Acoustics, Speech, Singnal Processing, vol. 1, pp. 304-307.
[15] Kuhn, R., Junqua, J.-C., Nguyen, P. and Niedzielski, N. 2000. "Rapid speaker adaptation in eigenvoice space." IEEE Trans. Speech and Audio Proc., vol. 8, no. 6, pp. 695-707.
[16] Kuhn, R. Perronnin, F. and Junqua, J.-C. 2001. "Time is money: why very rapid adaptation matters." in Proc. Adaptation Methods for Speech Recognition, ISCA ITR-Workshop, Sophia-Antipolis, France, pp. 33-36.
[17] Lee, C.-H. 1997. "On stochastic feature and model compensation approaches to robust speech recognition." Speech Comm., vol. 25, pp. 29-47.
[18] Lee, C.-H. and Huo, Q. 2000. "On adaptive decision rules and decision parameter adaptation for automatic speech recognition." Proc. of IEEE, vol. 88, no. 8, pp. 1241-1269.
[19] Leggetter, C. J. and Woodland, P. C. 1995. "Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models." Computer Speech and Language, vol. 9, pp. 171-185.
[20] Rubin, D. and Thayer, D. 1982. "EM algorithms for factor analysis." Psychometrika, vol. 47, pp. 69-76.
[21] Tipping, M. and Bishop, C. 1999. "Mixtures of probabilistic principal component analysers." Neural Computation, vol. 11, no. 2, pp. 443-482.
[22] Woodland, P. C. 2001. "Speaker adaptation for continuous density HMMs; a review." in Proc. Adaptation Methods for Speech Recognition, ISCA ITR-Workshop, Sophia-Antipolis, France, pp. 11-19.

접수일자: 2004. 08. 25
게재결정: 2004. 11. 24

▲ 김동국
광주광역시 북구 용봉동 300번지 (우: 500-757)
전남대학교 전자컴퓨터정보통신 공학부
Tel: +82-62-530-1794, Fax: +82-62-530-1750
E-mail: dkim@chonnam.ac.kr

▲ 장준혁
Eng-1, Department of Electrical and Computer
University of California, Santa Barbara,
Tel: +1-805-685-0753
E-mail: jhchang@ece.ucsb.edu

▲ 김남수
서울시 관악구 신림동 산 56-1 (우: 151-742)
서울대학교 전기컴퓨터공학부
Tel: +82-2-840-8419 (O), Fax: +82-2-840-8419
E-mail: nkim@snu.ac.kr