http://dx.doi.org/10.14400/JDC.2015.13.1.243

Bayesian 기법의 모수 추정을 이용한 결정트리 상태 공유 모델링

오상엽
가천대학교 컴퓨터미디어융합학과

Decision Tree State Tying Modeling Using Parameter Estimation of Bayesian Method

SangYeob Oh
Dept. of Computer Media Convergence, Gachon University

요 약 인식 모델을 구성할 때 정의되지 않은 모델, 인식 모델 구성 후에 추가되어진 모델, 모델이 부족하여 하나의 모델 클러스터링으로 모델링하여 생성된 인식 모델들은 인식률 저하의 원인이 된다. 이러한 원인을 개선하기 위하여 Bayesian 기법의 모수 추정을 이용한 결정트리 상태 공유 모델링 방법을 제안하였다. 제안 방법은 Bayesian 기법의 파라미터 추정을 통하여 탐색된 결과로부터 결정트리 기반 상태 공유 모델링의 최대 확률 기법에 따라 인식 모델을 결정한다. 본 논문에서 제안하여 시뮬레이션 데이터를 이용한 실험 결과에서 제안한 군집화 방식을 비교하여 1.29%의 음성인식 오류감소율을 보였으며, 기존 군집화 방식에 비해 개선된 성능을 보였다.

주제어 : HMM, 어휘 인식, Bayesian, 결정 트리, 공유 모델링

Abstract Recognition model is not defined when you configure a model, Been added to the model after model building awareness, Model a model of the clustering due to lack of recognition models are generated by modeling is causes the degradation of the recognition rate. In order to improve decision tree state tying modeling using parameter estimation of Bayesian method. The parameter estimation method is proposed Bayesian method to navigate through the model from the results of the decision tree based on the tying state according to the maximum probability method to determine the recognition model. According to our experiments on the simulation data generated by adding noise to clean speech, the proposed clustering method error rate reduction of 1.29% compared with baseline model, which is slightly better performance than the existing approach.

Key Words : HMM, Vocabulary Recognition, Bayesian, Decision Tree, Tying Modeling

Received 24 July 2014, Revised 30 August 2014
Accepted 20 January 2015
Corresponding Author: SangYeob Oh(The University of Gachon)
Email: syoh1234@gmail.com

ISSN: 1738-1916

© The Society of Digital Policy & Management. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서론

음성인식 알고리즘 개발에 있어 기존 연구들에서는 퍼지로직, Neural Network 및 HMM(Hidden Markov Model) 방법들을 주로 사용하고 있으며, 인식률을 높이기 위한 다양한 방법 및 알고리즘 개발이 수행되고 있다. HMM은 시공간적인 정보를 통한 모델링과 학습 및 인식을 위한 효과적이고 우수한 알고리즘을 가지고 있어 여러 분야에서 응용되고 있으며, 음성인식에 있어 가장 널리 사용되어지고 있다[1].

음성 인식을 위한 모델 파라미터들로 부터 데이터 부족 문제가 발생하게 되는데 이는 인식을 위한 모델별 훈련용 데이터의 량이 일정하지 않기 때문이다. 따라서 모델의 가용성과 모델의 복잡성을 고려하여 균형을 유지할 수 있는 방법으로 공유 방법이 사용된다[2]. 인식 모델의 공유 방법을 이용하여 모델을 생성하면 최소한의 복잡성으로 데이터의 특성을 표현하는 일반적인 모델을 구성할 수 있다. 인식 모델의 공유 방법은 파라미터의 수를 줄이므로 인식에 필요한 계산량을 줄일 수 있고 모델의 효율적인 훈련을 가능하게 한다. 인식 모델 공유 방법은 공유 수준에 따라 모델 공유, 상태 공유로 나눌 수 있다[3].

인식 모델 공유 방법은 모델들을 대상으로 하여 파라미터들을 공유할 수 있는 집합들을 결정하며 이러한 집합들을 사용함으로써 보다 강인한 모델을 추정하며 군집화에 사용될 모델 파라미터들의 초기 추정치를 생성하기 위하여 각 문맥들에 대한 데이터가 반드시 필요하다[4]. 인식 모델을 구성할 때 정의되지 않은 모델과 인식 모델 구성 후에 추가되어진 모델로 인식을 수행할 경우 인식할 모델이 없거나 다른 모델로 인식하여 인식률이 저하되는 문제점이 발생한다. 또한 모델이 부족하여 하나의 모델 클러스터링으로 모델링하여 생성된 인식 모델은 정확성을 확보하지 않아 인식률 저하의 원인이 된다. 이러한 문제점을 개선하기 위하여 Bayesian 기법의 모수 추정을 이용한 결정트리 상태 공유 모델링 방법을 제안하였다. 제안 방법은 음성으로부터 얻은 파라미터에 대한 정보와 파라미터에 대한 과거의 경험 또는 주관을 사전 분포로 표현하여 모델을 구성하여 보다 정확한 파라미터의 불확실성에 결정트리를 이용한 상태 공유 모델링을 수행한다. Bayesian 기법의 파라미터 추정을 통하여 탐색된 결과로부터 결정트리 기반 상태공유 모델링의 최대 확률 기법에 따라 인식 모델을 결정한다.

본 논문에서 제안하여 시뮬레이션 데이터를 이용한 실험 결과에서 제안한 군집화 방식을 비교하여 1.29%의 음성인식 오류감소율을 보였으며, 기존 군집화 방식에 비해 조금 개선된 성능을 보였다. 제2장에서는 HMM과 모델 군집화를 설명하고, 제3장에서는 본 논문에서 제안한 시스템에 대하여 설명한다. 제4장에서는 제안한 시스템의 실험결과에 대하여 설명하고 제5장에서 결론을 기술한다.

2. 관련 연구

2.1 HMM 모델

파라미터의 패턴들을 적용하여 모델링, 학습 및 음성인식을 수행하기 위한 방법으로 HMM을 사용한다. 또한 입력과 근사치의 출력을 생성하는 은닉 상태 집합의 경로를 검출하는 Viterbi 알고리즘을 사용하여 최적 확률의 인식을 수행한다. HMM은 관측할 수 없는 “hidden” 과정와 음성 신호로부터 “hidden” 과정의 상태로 유도되는 음향학적 벡터를 연결하는 관측 과정으로 구성된다. HMM에서는 관측할 수 없는 음성의 통계적인 특성을 관측 가능한 벡터열을 통해 추정함으로서 음성의 통계적인 변이성을 반영한다[5].

HMM 모델에서 상태열 q에 대한 관측열의 확률은 다음 식(1)로 표현된다.

$$\begin{aligned} P(O|q,\lambda) &= \prod_{t=1}^{T} P(o_t|q_t,\lambda) \\ &= \sum_{q\in Q} \pi_{q_1} b_{q_1}(o_1) a_{q_1 q_2} b_{q_2}(o_2) \cdots a_{q_{T-1} q_T} b_{q_T}(o_T) \\ &= \sum_{q\in Q} \prod_{t=1}^{T} a_{q_{t-1} q_t} b_{q_t}(o_t) \end{aligned} \quad (1)$$

HMM 알고리즘은 관측이 불가능한 프로세스를 관측이 가능한 다른 프로세스를 통해 추정하는 이중 확률 처리 기법을 사용하여 모델링한다. 특히 인식 가능한 기준 패턴을 음소 단위와 음절 단위로 설정하여 모델을 구성하고 인식하는 방법을 사용한다[6].

2.2 군집화

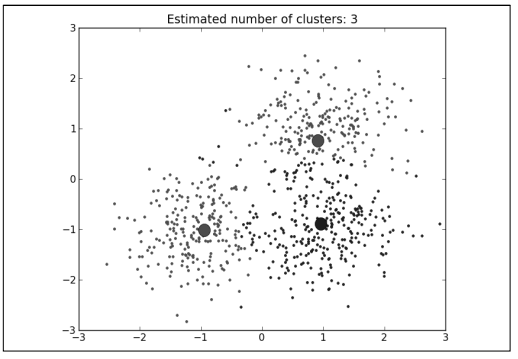
데이터를 구성하고 있는 관측대상을 어떤 기준에 의

하여 분류하는 통계기법이다. 패턴 인식 및 기계학습에서 중요한 기술 중의 하나인 군집화 방법에는 k-means 기법이 있다. k-means 기법은 표본점들을 거리에 따라 군집으로 묶는 역할을 한다[7]. 변수로 구성된 모델 클러스터링(Clustering)을 위해 사용하며 계산 속도가 빠르고 대량의 자료를 군집화하는데 효과적이다.

객체(Object)들은 클러스터링의 대상이 되며 특성을 나타내는 속성을 지닌다. 따라서 클러스터링을 통해서 특정 군집에 나뉘게 되며 각 군집에 소속된 객체들은 속성 정보를 소유하게 된다. 객체에 대한 클러스터링 결과를 분석하면 각 군집에 분포된 객체들의 분포도에 대한 정보를 얻을 수 있으며, 가장 가까운 중심점을 갖는 군집에 각 항목을 할당하는 과정을 반복하여 k개의 군집으로 항목들을 나뉘게 된다[8]. 거리 기반 클러스터링 방법으로 HMM 모델을 다차원 공간상의 점으로 표시하고 거리를 계산하여 전체 HMM 모델들의 집합을 k개의 군집으로 나눈다. $d(a,k)$는 대상 a와 k사이의 거리를 나타내고, X_{ij}는 대상 j의 변수를 나타내며, i는 좌표를 나타낸다. n은 측정 변수의 개수를 의미하며 다음 식(2)로 계산된다.

$$d(a,k) = \sqrt{\sum_{i=1}^{n}(X_{Ai} - X_{Bi})^2} \tag{2}$$

위의 식을 이용한 거리 기반 클러스터링의 k-means 기법을 이용한 클러스터링의 결과를 그림 1에 나타내었다.



[Fig. 1] K-means implementation

3. 모수 추정을 이용한 상태 공유 모델링

3.1 베이시안 기법

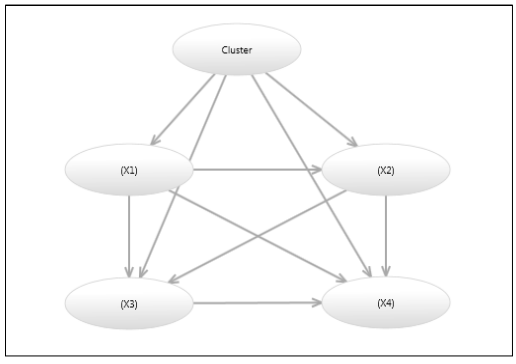
특정 확률밀도함수의 모수를 추정하기 위해서는 다양한 방법이 적용될 수 있지만 본 연구에서는 Bayesian 기법을 사용하였다. Bayesian 기법은 확률적인 접근론에 기반하기 때문에 모형의 모수는 임의계수로 취급된다. 따라서 고전모형에서는 모형의 모수가 점추정이었으나 Bayesian 기법에서는 확률분포로 추정하며 알려진 모델을 이용하여 알지 못하는 모델에 대한 학습을 반영하는 과정으로 수행된다. Bayesian의 정리를 연속 확률밀도함수로 나타내면 Bayesian의 정리는 다음 식(3)과 같이 표현된다[9].

$$p(\theta|x, x_{n+1}) \propto p(x_{n+1}|\theta) \bullet p(\theta|x) \tag{3}$$

θ는 추정해야 하는 모수를 나타내고, x는 알고 있는 모델을 나타낸다. $p(\theta|x)$는 사후 분포(posterior distribution)이며, $p(x|\theta)$는 주어진 모델 모수의 데이터의 함수인 우도 함수(likelihood function)이다.

Bayesian 기법을 이용한 모수의 추정은 모델로부터 얻은 모수에 대한 정보와 모수에 대한 과거의 경험 또는 주관을 사전 분포로 표현함으로써 보다 정확한 모수의 불확실성에 대한 추정을 수행한다. Bayesian 모수 추정 기법을 그림 2에 나타내었다.

Metropolis-Hastings 알고리즘을 수행하고 최종적으로 몬테카를로 적분을 적용하면 추정된 모수들의 평균과 같은 통계적 특성치를 근사적으로 얻게 된다[10].



[Fig. 2] Bayesian Parameter Estimation

또한 기존 통계학에서 사용되는 신뢰 구간(confidence interval)과 유사한 개념의 신용 구간(credible interval)을 구하여 추정된 모수의 불확실성을 표현한다. 이를 반복하여 추출된 모수의 개수 n과 유의수준 a에서 Bayesian은 다음 식(4)와 같이 나타낸다.

$$\tau = \frac{\sum_{i<j} n(x_i - x_j)n(y_i - y_j)}{\sqrt{\left(\frac{\Sigma t_i(t_i-1)}{2}\right)\left(\frac{\Sigma u_i(u_i-1)}{2}\right)}} \quad (4)$$

n은 변수의 관측치이며, t_i는 주어진 순위에서 동일한 X의 관측치이고, u_i는 주어진 순위에서 동일한 Y의 관측치이다. $n(z)=1$ 이면 $z>0$, $n(z)=0$이면 $z=0$, $n(z)=-1$ 이면 $z<0$를 나타낸다. 계산된 계수는 정규화를 통해 0~1사이의 값을 갖게 되고 추정된 결과를 나타낸다.

3.2 공유 모델링

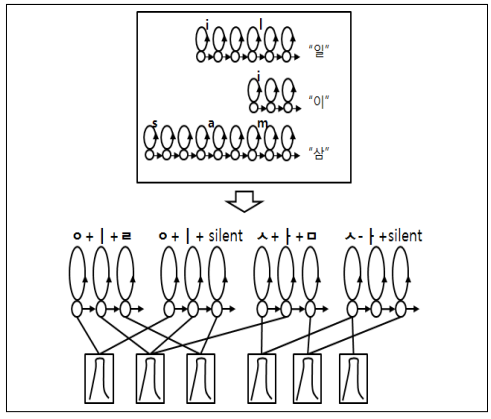
HMM으로 구성된 모델에 대해 상태들의 집합을 S로 나타내고 $L(S)$를 S의 log likelihood로 나타내면 집합 S에 있는 모델의 대한 모든 상태들이 3-states로 묶여 인식 모델 훈련 데이터의 프레임 집합인 F로 생성된다. 집합 F에서 3-states로 묶인 모델들에 대해 공통 평균 $\mu(S)$과 분산 $\Sigma(S)$를 공유하게 된다. 또한 묶인 상태들의 상태별 프레임의 정렬이 바뀌지 않은 상태에서 $L(S)$는 다음 식(5)와 같이 나타낸다.

$$L(S) = \sum_{f=F}\sum_{s=S} \log(\Pr(o_f;\mu(S)))^{*} G \quad (5)$$

G는 상태 s에 의해 생성되는 사후 확률을 나타내고 O_f는 관측 프레임을 나타낸다. 중심 음소로 갖는 HMM의 확률 분포들을 모아서 하나의 커다란 확률 풀을 만들어 각 트라이폰의 관측 확률의 공통 확률 풀에 속한 모델을 사용한다. 또한 확률의 가중치를 다르게 반영하여 나타난 모델 특징 벡터를 모아서 군집화한다.

공유 구조를 이용하여 모델들을 생성하기 위해서는 최소한의 복잡성으로 데이터의 특성을 유지하는 모델을 구성하는 것이다. 공유 구조를 사용하는 이유는 증가하는 파라미터를 줄일 수 있고 모델들을 인식하기 위한 훈련의 효율성을 증대시킬 수 있다. 인식을 위한 모델 파라미터의 수를 줄인다는 것은 인식에 필요한 계산량을 줄일 수 있기 때문이다. 공유 모델링에 대한 절차는 다음과 같이 진행되며, 그림 3에 상태를 공유하여 군집화하는 형태를 나타내었다.

1) 모델 객체에 대한 클러스터링 결과를 분석하여 각 군집에 분포된 객체들의 분포도에 대한 정보를 확인한다.
2) 가장 가까운 중심점을 갖는 군집에 각 항목을 할당하는 과정을 반복하여 k개의 군집으로 항목들을 나뉘게 된다.
3) 거리 기반 클러스터링 방법으로 HMM 모델을 다차원 공간상의 점으로 표시하고 거리를 계산하여 전체 HMM 모델들의 집합을 k개의 군집으로 나눈다.



[Fig. 3] Cluster tied-state

4. 실험 결과

본 논문에서 제안한 Bayesian 기법의 모수 추정을 이용한 결정트리 상태 공유 모델링 방법의 성능 검증을 위하여 어휘 인식 실험을 수행하였다. 음성인식을 위한 음향모델은 트라이폰 단위의 HMM을 사용하고 트리 기반 군집화를 이용하여 상태 tying된 3 상태의 left-to-right HMM을 사용하였다. 훈련 과정과 실험 환경과의 불일치 문제를 해결하기 위해 잡음 처리는 워너 필터를 사용하였으며 음성 인식 목록은 지하철역명 30개로 구성하였다.

인식 실험에서는 실험에 참가한 화자가 음성 인식 목록을 3회 발음하여 총 100단어를 대상으로 실험을 수행하였다[11].

제안한 시스템의 성능 평가를 위하여 군집화 방식에서 노드 수가 3개일 때 각 군집별 음향 모델을 군집 데이터를 이용하여 재훈련한 경우와 군집 데이터를 적응 데이터로 이용하여 적응한 경우의 단어오류율(WER:Word Error Rate)[12]을 비교하였다. 전체 훈련 데이터가 충분하지 않은 경우, 감소된 군집별 데이터로 인한 성능 저하는 적응 과정을 통하여 오류감소율(ERR:Error Reduction Rate)[13]로 성능 개선이 이루어졌음을 확인할 수 있었다.

기존 군집화 방식을 시뮬레이션 DB에 적용하고, 결정트리구조의 3개 군집을 사용한 인식 결과를 표 1에 나타냈으며, 각 군집별로 단어오류율과 오류감소율을 구하였다. 군집별 성능이 모든 군집들에서 베이스라인 성능보다 우수하였고 전체적으로 베이스라인 성능에 비해서 약 9.70%의 오류감소율을 얻었다.

〈Table 1〉 Speech Recognition Performance for Each Cluster in the Existing

Node	Date size	WER(%)		ERR(%)
		Baseline	Clustering	
0	576	9.37	9.37	
00	191	14.67	13.17	8.39
01	1096	9.12	8.12	11.00
Total	1,863	11.05	10.22	9.70

군집별 데이터 분포 특성에서 노드 0은 SNR에 따라 노드 00과 노드 01로 분할되는 경향이 있는데, 노드 00에는 SNR이 높은 데이터가 많이 분포하며, 노드 01에는 SNR이 낮은 데이터가 많이 분포하였다. SNR이 높을수록 분할이 잘 되었다.

<Table 2>는 제안한 Bayesian 기법의 모수 추정을 이용한 결정트리 상태 공유 모델링 방법을 사용했을 때의 음성인식 성능을 나타내었다. 적응 평균을 이용한 군집화의 경우, 베이스라인 대비 약 8.41%의 오류감소율로 성능 개선이 이루어졌음을 확인하였다. 이는 기존 상태 모델링 방식보다 오류감소율의 측면에서 1.29% 정도 우수하다. 군집별 데이터 분포 특성에서는 노드 0에서의 SNR에 따른 분할이 기존 방식보다 잘 이루어진 것을 관찰할 수 있다. SNR이 낮은 데이터들이 주로 분포하고 있는 노드 01은 낮은 SNR의 데이터들 위주로 군집화 되어 있다.

〈Table 2〉 Speech Recognition Performance for Each Cluster in the proposed clustering method

Node	Date size	WER(%)		ERR(%)
		Baseline	Clustering	
0	637	10.99	10.99	
00	896	4.01	3.12	6.59
01	576	12.12	11.07	10.23
Total	2,109	9.04	8.39	8.41

제안한 Bayesian 기법의 모수 추정을 이용한 결정트리 상태 공유 모델링 방법은 기존 상태 모델링 방식에 비해서 SNR 및 군집화가 잘 이루어지는 것을 표 1과 표 2에서 각 군집별 베이스라인 인식 성능의 편차 정도를 통해서 확인하였다. 군집화가 잘 될수록 이를 이용한 음성인식 성능이 향상될 수 있지만, 군집 선정이 정확하지 않다면 성능 향상을 보장할 수 없다.

5. 결론

기존 Bayesian 기법의 모수 추정을 사용하여 데이터 분석 방법을 본 논문에서는 결정트리 상태 공유 모델링 방법에 적용하여 제안하였다. 음성으로부터 얻은 파라미터에 대한 정보와 파라미터에 대한 과거의 경험 또는 주관을 사전 분포로 표현하여 모델을 구성하는 Bayesian 기법의 모수 추정을 정확한 파라미터을 이용하여 결정트리를 구성하고 상태 공유 모델링을 수행하였다. Bayesian 기법의 파라미터 추정을 통하여 탐색된 결과로부터 결정트리 기반 상태공유 모델링의 최대 확률 기법에 따라 인식 모델을 결정하여 인식 결과를 도출한다. 따라서 본 논문은 음성 인식 성능 저하를 극복하기 위해 제안된 Bayesian 기법의 모수 추정을 이용한 결정트리 상태 공유 모델링 방법을 제안하고 성능 실험을 하였다. 시뮬레이션 데이터를 이용한 실험 결과에서 제안한 군집화

방식을 비교하여 1.29%의 음성인식 오류감소율을 보였으며, 기존 군집화 방식에 비해 개선된 성능을 보였다.

향후 연구 과제로는 음성 데이터 사이즈를 일률적으로 구성하여 실험하는 방법과 지하철 역명의 데이터를 확대하여 실험하고 데이터가 확대되었을 경우 인식률에 미치는 영향을 확인하여 인식률을 향상시킬 수 있는 연구가 필요하다.

REFERENCES

[1] A. Srinivasan, Speech Recognition Using Hidden Markov Model, Applied Mathematical Sciences, vol. 5, no. 79, pp. 3943-3948, 2011.

[2] Chan-Shik Ahn, Sang-Yeob Oh. Gaussian Model Optimization using Configuration Thread Control In CHMM Vocabulary Recognition. The Journal of Digital Policy and Management. Vol. 10, No. 7, pp. 167-172, 2012.

[3] Chan-Shik Ahn, Sang-Yeob Oh. Echo Noise Robust HMM Learning Model using Average Estimator LMS Algorithm. The Journal of Digital Policy and Management. Vol. 10, No. 10, pp. 277-282, 2012.

[4] Chan-Shik Ahn, Sang-Yeob Oh. CHMM Modeling using LMS Algorithm for Continuous Speech Recognition Improvement. The Journal of digital policy and management. Vol. 10, No. 11, pp. 377-382, 2012.

[5] Chan-Shik Ahn, Sang-Yeob Oh. Vocabulary Recognition Retrieval Optimized System using MLHF Model . Journal of the Korea Society of Computer and Information. Vol. 14, No. 10, pp. 217-223, 2009.

[6] Beaufays, F., Vanhoucke, V., & Strope, B. Unsupervised discovery and training of maximally dissimilar cluster models. Proc. Interspeech, pp. 66-69, 2010.

[7] Zhang, Y., Xu, J., Yan, Z. J., & Huo, Q. An i-vector based approach to training data clustering for improved speech recognition. Proc. Interspeech, pp. 1247-1250. 2011.

[8] Tsao, Y. & Lee, C. H. An ensemble speaker and speaking environment modeling approach to robust speech recognition. IEEE Trans. Audio, Speech, and Language Processing, Vol. 17, No. 5, pp. 1025-1037, 2009.

[9] Sang-Yeob Oh. Improving Phoneme Recognition based on Gaussian Model using Bhattacharyya Distance Measurement Method. Journal of Korea Multimedia Society. Vol. 14, No. 1, pp. 85-93, 2011.

[10] Campbell, W. M., Sturim, D. E., Reynolds, D. A., Solomonoff, A. SVM based speaker verification using a GMM supervector kernel and NAP variability compensation. Proc. ICASSP, No. 1, pp. 97-100, 2006.

[11] Sang-Yeob Oh. Selective Speech Feature Extraction using Channel Similarity in CHMM Vocabulary Recognition. The Journal of digital policy and management. Vol. 11, No. 10, pp. 453-458, 2013.

[12] Ban, S. M., Kang, B. O., Lee, Y. K., Kim, H. S. Automatic clustering of speech data using the distance between the cepstral mean vectors. Proc. 2012 Fall Conf. of the Korean Society of Speech Sciences, pp. -36, 2012.

[13] Lee, S. J., Kang, B. O., Jung, H. Y., Lee, Y. K. Kim, H. S. Statistical model-based noise reduction approach for car interior applications to speech recognition. ETRI Journal, Vol. 32, No. 5, pp. 801-809, 2010.

오 상 엽(Oh, Sang Yeob)

・1991년 2월 : 광운대학교 대학원 전자계산학과 (이학석사)
・1999년 2월 : 광운대학교 대학원 전자계산학과 (이학박사)
・2007년 2월 ~ 현재 : 가천대학교 IT대학 인터랙티브미디어학과 교수
・관심분야 : 버전관리, 형상관리, 음성/음향 신호 처리, 차량 통신
・E-Mail : syoh1234@gmail.com