The Journal of Digital Policy & Management 2013 Sep; 11(9): 153-158

어휘 인식 시스템에서 학습 모델 분류를 위한 결정 트리 학습 알고리즘

오상엽*
가천대학교 글로벌캠퍼스 IT대학 컴퓨터미디어융합학과*

Decision Tree Learning Algorithms for Learning Model Classification in the Vocabulary Recognition System

Sang-Yeob Oh*
Dept. of Computer Media Convergence, College of IT, Gachon University*

요 약 인식 대상 학습 모델이 분류되어 있지 않거나 명확하게 분류되지 않은 경우 어휘 인식을 결정하지 못하여 인식률이 저하되며 학습 모델 분류 형태가 변경되거나 새로운 학습 모델이 추가되면 인식 모델의 결정 트리 구조가 변경되어야 하는 구조적 문제가 발생한다. 이러한 문제점을 해결하기 위하여 학습 모델 분류를 위한 결정 트리 학습 알고리즘을 제안한다. 음운 현상이 충분히 반영된 음성 데이터베이스를 구성하고 학습 효과를 확보하기 위하여 학습 모델 분류를 위한 결정 트리 방법을 사용하였다. 본 연구에서는 실내 환경에 대하여 어휘 종속 인식과 어휘 독립 인식 실험을 수행한 결과 실내 환경의 어휘 종속 실험에서는 98.3%의 인식 성능을 보였고, 어휘 독립 실험에서 98.4%의 인식 성능을 보였다.

주제어 : 음성 인식, 특징 추출, 학습 모델, 결정 트리, 학습 알고리즘

Abstract Target learning model is not recognized in this category or not classified clearly failed to determine if the vocabulary recognition is reduced. Form of classification learning model is changed or a new learning model is added to the recognition decision tree structure of the model should be changed to a structural problem. In order to solve these problems, a decision tree learning model for classification learning algorithm is proposed. Phonological phenomenon reflected sound enough to configure the database to ensure learning a decision tree learning model for classifying method was used. In this study, the indoor environment-dependent recognition and vocabulary words for the experimental results independent recognition vocabulary of the indoor environment-dependent recognition performance of 98.3% in the experiment showed, vocabulary independent recognition performance of 98.4% in the experiment shown.

Key Words : Speech Recognition, Feature Extraction, Learning Model, Decision Tree, Learning Algorithm

* 이 논문은 2013년도 가천대학교 교내연구비 지원에 의한 결과임.(GCU-2013-R187)
Received 8 July 2013, Revised 3 August 2013
Accepted 20 September 2013
Corresponding Author: SangYeob Oh(The University of Gachon)
Email: syoh1234@gmail.com

ISSN: 1738-1916

1. 서론

멀티미디어 데이터의 증가로 인하여 데이터를 의미 있는 카테고리로 분류하기 위한 자동 인식 및 분류 기술이 새로운 연구 분야로 나타나고 있다. 신호 처리 기술 및 알고리즘의 발전에 힘입어 인식 시스템의 결과들을 활용하는 연구가 활발하게 진행되고 있다[1].

어휘 인식 시스템은 인식 대상 어휘를 선정하여 발성 데이터를 수집하고 음성 데이터베이스를 구축한다. 학습 모델을 구성하기 위하여 인식할 대상 어휘에 대한 음소와 유사 음소를 생성한다. 어휘 인식 시스템은 인식 대상 어휘에 종속되지 않는 결정 트리 방법을 사용하며 어휘에 대한 음운 현상을 충분히 반영하여 음향 모델을 통하여 단위 모델로 학습하고 인식 어휘가 추가 및 변경되어도 인식할 수 있도록 한다. 하지만 학습 모델을 인식함으로 이미 생성된 학습 모델에 대해서는 인식 성능이 우수하게 나타나지만 생성되지 않은 학습 모델에 대해서는 인식률이 떨어지게 된다. 이러한 인식률 저하는 학습 모델의 생성과 생성된 모델의 분류가 명확하지 않기 때문에 나타난다[2,3].

인식 대상 학습 모델이 분류되어 있지 않거나 명확하게 분류되지 않은 경우 어휘 인식을 결정하지 못하여 인식률이 저하되며 학습 모델 분류 형태가 변경되거나 새로운 학습 모델이 추가되면 인식 모델의 결정 트리 구조가 변경되어야 하는 구조적 문제가 발생한다. 이러한 문제점을 해결하기 위하여 학습 모델 분류를 위한 결정 트리 학습 알고리즘을 제안한다.

음운 현상이 충분히 반영된 음성 데이터베이스를 구성하고 학습 효과를 확보하기 위하여 학습 모델 분류를 위한 결정 트리 방법을 사용하였다. 학습 모델 분류를 위한 결정 트리 방법은 음운 현상을 반영한 음향 모델을 단위 모델로 학습하고 인식 어휘의 추가 및 변경 작업을 용이하게 처리한다. 문맥 종속적인 음소가 추가되었을 때 루트노드에서부터 현재의 문맥에 관한 질문의 결과에 따라 다음 노드를 선택하는 방식의 트리 순회를 통하여 하나의 모델을 선택하게 함으로써 결정 트리 학습 알고리즘을 구성하였다.

본 연구에서는 실내 환경에 대하여 어휘 종속 인식과 어휘 독립 인식 실험을 수행한 결과 실내 환경의 어휘 종속 실험에서는 98.3%의 인식 성능을 보였고, 어휘 독립 실험에서 98.4%의 인식 성능을 보였다.

본 논문의 구성은 제 2장에서는 어휘 인식 시스템에 대해 간략히 소개하고, 제 3장에서는 본 논문에서 제안한 학습 모델 분류를 위한 결정 트리 학습 알고리즘에 대하여 설명한다. 제 4장에서는 제안한 시스템의 실험 결과에 대하여 설명하고 제 5장에서 결론을 맺는다.

2. 관련 연구

2.1 학습 모델

학습 모델은 주어진 표본 데이터 집합의 분포 밀도를 단 하나의 확률 밀도 함수로 모델링하는 방법을 개선한 밀도 추정 방법으로 복수 개의 확률 밀도 함수로 데이터의 분포를 모델링하는 방법이다. 단일 특징으로는 모델링할 수 없는 복수 개의 중심점을 가지는 1차원 데이터와 2차원 환형 데이터에 대하여 견고하게 모델을 구성한다[4].

확률 밀도 함수는 특징 분포뿐 아니라 다른 분포가 될 수도 있다. 특징 최적화 밀도는 단지 확률 밀도 함수를 특징 분포로 가정하는 경우이다. 결국 최종적인 전체 확률 밀도 함수는 M개의 특징 확률 밀도 함수의 선형 결합으로 다음 식과 같이 나타낸다.

$$p(x|\theta)=\sum_{i=1}^{M} p(x|\omega_i,\theta_i)P(\omega_i) \tag{1}$$

$p(x|\omega_i,\theta_i)$는 데이터 x에 대하여 ω_i번째 성분 파라미터 θ_i로 이루어진 확률 밀도 함수를 의미하며, $P(\omega_i)$는 혼합 가중치로 각 확률 밀도 함수의 상대적인 중요도를 의미한다. 혼합 가중치를 사전 확률과 같은 형태로 α_i라고 하면 다음 식과 같은 제약조건이 따른다.

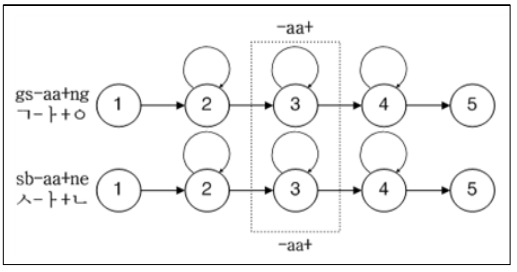
$$0 \le \alpha_i \le 1,\ \sum_{i=1}^{M} \alpha_i = 1 \tag{2}$$

확률 밀도 함수가 특징 분포를 따를 경우 θ_i는 다음 식과 같은 파라미터 집합으로 나타난다.

$$\theta_i = \begin{pmatrix} \mu_1, \mu_2, ..., \mu_M, \\ \sigma_1, \sigma_2, ..., \sigma_M, \\ \alpha_1, \alpha_2, ..., \alpha_M \end{pmatrix} \quad (3)$$

전체 모델을 이루는 각 특징 성분은 완전 대각 또는 정방형 공분산 행렬의 형태를 나타내며 혼합 성분의 개수는 학습 데이터 집합의 크기에 따라 조절된다[5].

단일 특징 출력 확률 밀도 함수를 갖는 3상태 단 음소 모델 초기 집합을 생성하고 학습한다. 특징 벡터들을 추출하여 특징 값을 추정하여 구성한 특징 학습 모델을 그림 1에 나타내었다.



[Fig. 1] Feature Learning model

2.2 Haar-like feature와 Adaboost 분류기

Haar-like feature는 인식, 학습, 분류 등에 사용되며 사각형의 집합으로 구성된 특징 벡터의 집합으로써 신호 영역 누적 테이블을 이용하므로 빠른 시간에 특징의 적합도를 구하기 때문에 많이 사용되고 있다. 기초적인 간단한 사각형들로 구성된 특징만으로는 복잡한 신호를 인식하기 어렵기 때문에 단순한 Haar-like feature의 집합을 각각의 가중치와 함께 합산함으로써 복잡한 신호에 대한 인식기를 구성하여 사용한다[6].

Adaboost는 실제로 여러 분류기나 학습기를 조합하여 하나의 메타 학습기를 만드는 기법으로 하나의 학습 방법에 의해 발생하는 약점을 보완할 수 있다. 하지만 여러 학습 및 분류기를 사용함으로써 시간적인 오버헤드가 크다. Adaboost는 단순하면서 약한 분류기를 조합하여 하나의 분류기를 생성할 때 사용된다[7].

Haar-like feature는 이런 Adaboost의 성격에 매우 적합하며 각각의 특징을 일종의 분류기로서 간주하고 이들의 가중치를 조정해 하나의 최종 분류기를 생성한다[8].

약 분류기에 의해 판단된 결과 값을 입력으로 간주할 경우 최종 분류기는 각각의 입력에 가중치를 곱한 값들의 합산으로 생각할 수 있으며 이는 단층 신경망이나 SVM(Support Vector Machine)과 같은 성격을 지니게 된다. 이러한 방법은 비슷한 데이터이거나 정규분포로 구성된 그룹들의 분류에는 효율적이나 둘 이상의 종류로 나뉘어지는 그룹에 대해서는 구조상 표현이 어렵다[9].

비교적 유사한 특징들이 높은 가중치를 얻음으로써 자연스럽게 학습이 수행되지만 입력 데이터가 둘 이상의 유형으로 분류될 경우 상반된 특징간의 충돌로 인하여 적합한 학습 효과를 얻기 어렵다.

3. 결정 트리 학습 알고리즘

3.1 학습 모델 분류

분류기는 높은 성능과 정확도를 가지기 때문에 일반적인 패턴 분류를 위해서 주로 사용되는 방법이지만, 계산의 복잡성 때문에 높은 차원을 가진 특징이나, 방대한 양의 클래스를 분류 할 때는 적합하지 않다. 특징 벡터 차원의 크기에 덜 민감하고 분류 속도가 빠르며 정확한 분류 결과를 나타내는 것으로 알려진 랜덤 포레스트 분류기를 사용한다[10].

Input
- D_{target} : 트리의 최대 확장 깊이
- S_n: 모든 특징 샘플을 포함하는 학습 데이터
- 초기값 : $i=0, j=0, k=0, F_i=1$

(1) n개의 부스트랩 샘플을 S_n에 할당하고 특징 생성

(2) S_n으로부터 m개의 서브 부스트랩 샘플 선택

Loop : $D_k < D_{target}$ 또는 종료 조건 만족

$k=k+1$

① m개의 부스트랩을 사용하여 초기 트리 확장

② 각 내부 노드는 무작위 샘플로부터 p개의 특징을 선택하고 최적의 분할 함수를 결정.

③ 다른 p번째 특징을 사용하여 분할 함수 $f(v_p)$는 다음 수식에 의해 반복적으로 m개의 샘플 데이터를 left(I_l) 과 right(I_r)의 서브셋으로 분할

$$I_l = p \in I_n | f(v_p) < t,\ I_r = \frac{I_l}{I_n}$$

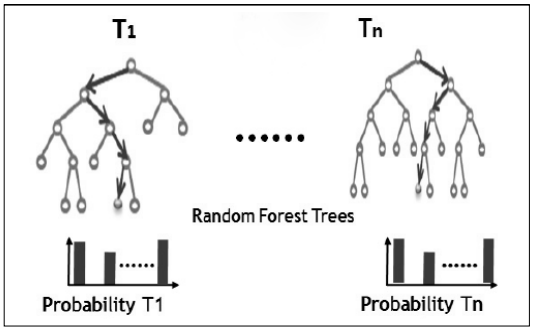
④ 임계값 t는 분할 함수 $f(v_p)$에 의해 범위 $t \in (\min_p f(v_p), \max_p f(v_p))$에서 랜덤하게 선택.

Loop ends

Breiman에 의해 제안된 랜덤 포레스트는 다수의 결정 이진 트리를 결합한 것으로, 각 이진트리에서는 랜덤한 방법으로 트리들을 성장시킨다[11,12]. 랜덤 포레스트는 결정 트리들을 기본으로 빠른 학습 속도와 많은 양의 데이터 처리 능력을 가지고 있다. 학습 데이터로 부터 각 특징들을 생성한 후, 랜덤 포레스트의 각 결정 트리는 아래와 같은 알고리즘에 의해 구축 되도록 설계하였다.

반복적인 학습은 information gain 이 0이거나 종단 노드가 최대 트리 깊이에 도달할 때 종료 된다.

그림 2와 같이 랜덤 포레스트의 각 트리의 구조는 이진이며, 하향식 형태를 갖는다.



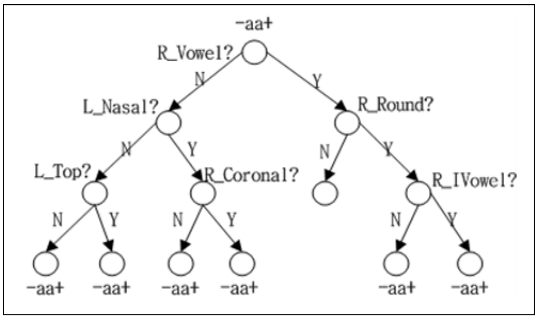
[Fig. 2] Speech Classification using Random Forest

3.2 결정 트리 학습 알고리즘

결정 트리는 분류 분야에서 널리 사용되는 방식으로 그림 2와 같이 각각의 특징 들을 결정 트리 형태로 구성하여 입력 값에 대한 판단을 수행하는 방식이다. 학습 과정 중에 엔트로피를 이용하여 더 효율적인 분류를 가능케 하는 특징 값을 상위 노드에 배치함으로써 작은 트리 구조로도 효율적으로 입력 값을 분류하는 분류기를 생성할 수 있다. 이는 XOR 문제에 대해서도 분류가 가능하며 복잡한 상태 공간에 대한 적응성이 뛰어나다[13].

모델 기반 공유는 모델이 제대로 훈련되지 않았거나 서로 닮은 경우에 비슷한 triphone들을 병합시켜 모델들을 더욱 많은 데이터로 사용하므로 시스템의 전체 크기는 줄이면서 더욱 정확한 모델을 생성한다. 모델 기반 공유는 왼쪽과 오른쪽의 문맥들을 독립적으로 처리할 수 없기 때문에 좌우의 문맥이 서로 다른 triphone들도 하나로 병합될 수 있는 mis-average 문제가 발생한다. 훈련 데이터베이스로부터 얻을 수 있는 모든 문맥 종속적인 모델을 생성하고 반복을 줄이기 위해 상태를 합병한다. 독립적으로 훈련된 모델들은 공유된 구조를 만들기 위해 비슷한 분포를 갖는 모델들로 클러스터링 된다[14].

공유 블록을 만든 후 결정 트리를 구성하며 그림 3과 같이 문맥 종속적인 모델에 대한 상태들의 결정을 통해 나온 최종 단 노드에 질의어에 대한 상태 공유 결과의 모델로 구성된다.



[Fig. 3] Decision Tree for Tri-phone / ㅏ / Value

많은 숫자의 단어 간 triphone들로 인하여 추정해야할 모델의 숫자가 많아지며, 그 가운데 상당수의 triphone들은 훈련데이터의 발생 횟수가 매우 적거나 전혀 발생하지 않는 현상이 일어나게 된다. 특별한 응용을 위해 필요한 triphone의 전체 개수는 음소 집합과 사전에 따라 결정하게 된다.

4. 실험 결과

본 논문에서 제안한 학습 모델 분류를 위한 결정 트리 학습 알고리즘의 성능 검증을 위하여 어휘 인식 실험을 수행하였다. 음성 인식 목록은 서울 시내의 지역명 50개, 지하철명 50개로 구성하였다.

제안한 시스템의 성능 평가를 위하여 기존 방식과 비교 실험을 하였다.

표 1과 2는 기존의 방식인 Euclidean[15], DTW[16] 그리고 제안한 방법을 실내 환경에서의 실험과 실외 환경에서의 실험을 나타낸다.

〈Table 1〉 lutdoor Environment Recognition Rate

Speech	DTW(%)	Euclidean(%)	Proposed Method(%)
Speech Dependent	96.3	97.1	97.5
	97.8	97.2	98.5
	97.5	97.6	98.9
Speech Independent	97.5	96.9	98.3
	97.2	96.5	98.5
	97.1	96.3	98.3

표 1은 실내 환경인 50~55dB에서 실험 하였으며 결과에서 보는 것과 같이 시스템 성능 평가 결과 어휘 종속 인식률은 Euclidean과 DTW는 각각 97.2%와 97.3%로 나타났고 제안한 방법은 98.3%로 나타났다. 어휘 독립 인식률에서도 Euclidean과 DTW는 각각 97.3%와 96.6%로 나타났고 제안한 방법은 98.4%로 나타났다.

〈Table 2〉 Outdoor Environment Recognition Rate

Speech	DTW(%)	Euclidean(%)	Proposed Method(%)
Speech Dependent	92.3	93.1	95.1
	92.9	92.8	95.8
	94.8	90.6	94.5
Speech Independent	91.6	96.1	97.3
	92.7	94.2	96.3
	91.8	93.2	93.7

표 2는 실외 환경은 70~75dB의 소음환경 하에서 실험하였으며 결과에서 보는 것과 같이 어휘 종속 인식률은 Euclidean과 DTW는 각각 93.3%와 92.2%로 나타났고 제안한 방법은 95.1%로 나타났다. 어휘 독립 인식률에서도 Euclidean과 DTW는 각각 92.0%와 94.5%로 나타났고 제안한 방법은 95.8%로 나타났다.

5. 결론

학습 모델을 인식함으로 이미 생성된 학습 모델에 대해서는 인식 성능이 우수하게 나타나지만 생성되지 않은 학습 모델에 대해서는 인식률이 떨어지게 된다. 이러한 인식률 저하는 학습 모델의 생성과 생성된 모델의 분류가 명확하지 않기 때문에 나타난다. 인식 대상 학습 모델이 분류되어 있지 않거나 명확하게 분류되지 않은 경우 어휘 인식을 결정하지 못하여 인식률이 저하되며 학습 모델 분류 형태가 변경되거나 새로운 학습 모델이 추가되면 인식 모델의 결정 트리 구조가 변경되어야 하는 구조적 문제가 발생한다. 이러한 문제점을 해결하기 위하여 학습 모델 분류를 위한 결정 트리 학습 알고리즘을 제안한다.

음운 현상이 충분히 반영된 음성 데이터베이스를 구성하고 학습 효과를 확보하기 위하여 학습 모델 분류를 위한 결정 트리 방법을 사용하였다. 본 연구에서는 실내 환경에 대하여 어휘 종속 인식과 어휘 독립 인식 실험을 수행한 결과 실내 환경의 어휘 종속 실험에서는 98.3%의 인식 성능을 보였고, 어휘 독립 실험에서 98.4%의 인식 성능을 보였다.

ACKNOWLEDGMENTS

This work was supported by the Gachon University research fund of 2013."(GCU-2013-R187)

REFERENCES

[1] 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.

[2] Jong-Young Ahn, Sang-Bum Kim, Su-Hoon Kim, Kang-In Hur. A study on Voice Recognition using Model Adaptation HMM for Mobile Environment. The Journal of the Institute of Webcasting, Internet and Telecommunication. Vol. 11, No. 3, pp. 175-179, 2011.

[3] Chan-Shik Ahn, Sang-Yeob Oh. Vocabulary Recognition Post-Processing System using Phoneme Similarity Error Correction. Journal of the Korea Society of Computer and Information. Vol. 15, No. 7, pp. 83-90, 2010.

[4] M. Cowling, R. Sitte. Comparison of techniques for environmental sound recognition. Pattern Recognition Letters, Vol. 24, No. 15, pp.

2895-2907, 2003.

[5] Yusuke Kida, Hiroyoshi Yamamoto. Minimum classification error interactive training for speaker Identification. IEEE International conference, Acoustic, Speech and Signal processing, Vol. 1, pp. 641-644, 2005.

[6] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In Proc. CVPR, pp.511 - 518, 2001.

[7] Quinlan, J.R. (1985b). Decision trees and multi-valued attributes. In J.E. Hayes & D. Michie (Eds.), Machine intelligence 11. Oxford University Press(in press).

[8] Ji-Eun Kim, In-Sung Lee. Speech/Mixed Content Signal Classification Based on GMM Using MFCC. Journal of the Institute of Electronics Engineers of Korea. Vol. 50, No. 2, pp. 185-192, 2013

[9] Ju-Hyun Kwak, Il-Young Woen, Chang-Hoon Lee. Learning Algorithm for Multiple Distribution Data using Haar-like Feature and Decision Tree. KIPS Transactions on Software and Data Engineering, Vol. 2, No. 1, pp. 43-48, 2013.

[10] V. Delaitre, I. Laptev, and J. Sivic. "Recognizing human action in still images: a study of bag-of-features and partial-based representations," in Proc. British Machine Vision Conf., pp.1-11, Wales, UK, Sep. 2010.

[11] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: spatial pyramid matching for recognizing natural scene categories," in Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition, pp. 2169-2178, NY, USA, Jun. 2006.

[12] M. Heikkilä, M. Pietikäinen, and C. Schmid, "Description of interest regions with local binary patterns," Pattern Recogn., vol.42, no. 3, pp. 425-436, 2009.

[13] Tariquzzaman, Md, Min, So-Hui, Kim, Jin-Yeong, Na, Seung-Yu. Modified HMM Decoder based on Observation Confidence for Speaker Identification. Proceedings of the Korean Institute of Intelligent Systems Conference. pp. 443-446. 2007.

[14] Y. G. Jiang C. W. Ngo, and J. Yang, "Towards optimal bag-of-features for object categorization and semantic video retrieval," in Proc. ACM Int. Conf. on Image and Video Retrieval, pp. 494-501, Amsterdam, Netherlands, 2007.

[15] 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.

[16] 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.

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