한국정보전자통신기술학회논문지(jkiiect)'17-02, Vol.10 No.1
논문 17-10-01-85 http://dx.doi.org/10.17661/jkiiect.2017.10.1.85

최대 빈도모델 탐색을 이용한 동물소리 인식용 소리모델생성

고유정*, 김윤중**

Sound Model Generation using Most Frequent Model Search for Recognizing Animal Vocalization

Youjung Ko*, Yoonjoong Kim**

요 약 본 논문에서는 동물소리 인식시스템을 위하여 최대 빈도모델 탐색 알고리즘을 고안하고 이를 이용한 소리모델을 생성하는 방법을 제안하였다. 소리모델 생성 방법은 동물종의 소리 데이터로부터 학습과정, 비터비 탐색과정 및 최대 빈도모델 탐색과정을 반복하면서 HMM(Hidden Makcov Model)모델의 구조(상태의 수와 GMM의 수)를 탐색하여 최적의 인식률을 갖는 모델집합이 생성하는 방법이다. 최대 빈도모델 탐색 알고리즘은 입력 소리 데이터를 비터비(Viterbi) 알고리즘으로 탐색하여 모델리스트를 생성하고 이 리스트 중에서 최대 빈도수의 모델을 탐색하여 최종 인식결과로 결정하는 방법이다. 알고리즘에서 소리특징으로 MFCC(Mel Frequency Cepstral Coefficient), 모델형식으로 HMM을 이용하고 C# 프로그래밍언어로 구현 하였다. 알고리즘의 성능을 평가하기 위하여 27종의 동물소리를 선정하고 실험을 하였으며 27개의 HMM 모델집합이 97.29 퍼센트의 인식률로 생성됨을 확인하였다.

Abstract In this paper, I proposed a sound model generation and a most frequent model search algorithm for recognizing animal vocalization. The sound model generation algorithm generates a optimal set of models through repeating processes such as the training process, the Viterbi Search process, and the most frequent model search process while adjusting HMM(Hidden Markov Model) structure to improve global recognition rate. The most frequent model search algorithm searches the list of models produced by Viterbi Search Algorithm for the most frequent model and makes it be the final decision of recognition process. It is implemented using MFCC(Mel Frequency Cepstral Coefficient) for the sound feature, HMM for the model, and C# programming language. To evaluate the algorithm, a set of animal sounds for 27 species were prepared and the experiment showed that the sound model generation algorithm generates 27 HMM models with 97.29 percent of recognition rate.

Key Words : Animal Vocalization Recognition, MFCC, Most Frequent Model Search Algorithm, HMM, Sound Model Generation

1. 서론

최근 멀티미디어 콘텐츠가 대량으로 생산되고 유통됨에 따라 신속하고 정확하게 콘텐츠의 특정 개체들을 검색할 수 있는 방법이 요구되고 있다. 오디오는 비디오에 비해 처리 정보량이 적고 사용방법이 간편하기 때문에 검색 시스템 제작의 질의어로 유용하게 사용할 수 있다. 질의 대상 콘텐츠는 상품정보 설명을 위한 오디오 일 수도 있고 영화장면 속의 특정 동물소리 또는 보안감시용 카메라의 녹음중인 비명소리일 수도 있다. 또한 최근 고성능의 이동단말기가 폭발적으로 보급됨에 따라 어떠한 질문도 즉시 해답을 얻을 수 있게 되었다. 그러나 미지의 동물소리에 대한 의문을

*Department of Computer Engineering, Hanbat National University
**Corresponding Author : Department of Computer Engineering, Hanbat National University
(yjkim@hanbat.ac.kr)
Received January 31, 2017 Revised February 04, 2017 Accepted February 04, 2017

갖게 되었을 때 인터넷으로 그 동물의 개체를 탐색하는 것은 쉽지 않다. 따라서 모바일 기기에 탑재 할 동물소리 인식 애플리케이션의 방법도 요구되고 있다. 이 애플리케이션이 보급되면 불특정 다수로부터 동물의 소리 획득이 가능하고 이를 빅데이터 방법으로 처리하여 동물의 서식지 분포, 이동경로 등 생태학적 이용에도 응용이 가능하다. 본 연구에서는 동물소리를 대상으로 동물을 식별하기 위하여 동물소리 인식시스템에서 핵심요소인 동물소리모델을 생성하는 방법과 종의 인식률을 제고하기 위한 최대 빈도모델 탐색알고리즘을 제안하였다.

동물소리 인식 시스템의 목적은 동물이 우는 소리(발성)를 이용하여 동물의 종을 인지하는 시스템을 개발하는 것이다. 동물의 발성은 종에 특화되도록 진화하였으므로 제각각 다른 발성주파수와 패턴을 가지고 있다. 따라서 동물의 종을 자동식별하기 위하여 동물의 발성을 이용하는 것은 자연스러운 일이고 이 인식기술은 환경 감시, 다양성 평가[1] 또는 동물원, 국립공원, 야생동물보호구역 및 동물농장 등에서 위험종의 출현을 탐지하는 매우 유용한 기술이다[2]

동물의 소리를 이용하여 종을 인식하기 위한 연구의 예로 연구[3] 에서는 조류, 고양이, 소 및 개의 소리를 포함하는 동물울음소리를 인식하기위하여 기계학습기술을 사용하였고, 연구[4]에서 16종의 동물소리 식별을 위하여 SVM(Support Vector Machine)을 채택하였다. 연구 [5]는 결정트리기술을 이용하여 야행성동물의 발성을 분류하고 MFCC(Mel Frequency Cepstral Coefficient) 특징을 사용하였다. 연구[6]은 웨이브릿을 이용하여 조류를 식별하고, 연구[7]에서는 MFCC 및 PNN(Probabilistic Neural Network)을 이용하여 곤충의 소리를 인식하였다. 이들 방법에서는 음성의 특징 및 분류방법으로 MFCC, SVM, PNN, 웨이블릿 등을 사용하고 있지만 종에 대한 모델을 정의하고 입력소리에 대하여 모델의 구조를 탐색하여 최적의 인식률을 추구하는 방식 및 음성인식기술을 사용하고 있지 않다. 또한 특정 소리에 특화된 인식기술의 연구이다. 본 방법에서는 인간음성 기술들(MFCC, HMM)을 이용하여 임의의 동물소리에도 적용가능하고 HMM의 구조를 자동탐색하여 최적으로 동물소리모델을 생성하는 알고리즘을 제안한다.

인간의 발성과 동물의 발성은 차이가 있다. 인간의 발성기관은 정보를 전달하기 위하여 복잡한 음향학적신호를 생성하도록 정밀하게 진화되었다. 다양하고 복합적인 주파수, 에너지 그리고 특정 패턴의 음향학적 신호를 생산한다. 이 신호들은 음소를 만들고 음소를 조합하여 의미가 포함되는 단어를 만들고 문장을 만들어 정보를 교환한다. 또한 같은 문장의 소리도 발음속도, 고저 및 반복 등의 음운정보를 추가하여 암시적 정보인 감성까지도 전달하기 위한 다양하고 정밀한 음성을 발성한다. 또한 이와 같이 발성된 음향학적신호로부터 음성을 인식하고 감성정보를 인식하는 다수의 연구들이 진행되어 왔다[8][9]. 동물소리의 인식에 이와 같이 풍부하게 연구되어진 음성인식기술을 적용하고 저 한다. 그러나 동물의 경우에는 발성기관의 단조로움으로 인하여 다양한 음향학적 신호를 생성하지 못한다. 따라서 다음과 같은 고려사항이 존재 한다. 동물의 소리는 특화된 신호를 생성하지만 그 종류가 단조로우므로 인간의 음성과 같이 음소단위로 정의하여 모델을 생성하는 것은 효율적이지 못하다. 따라서 한 종에 몇 개의 기본 모델로 정의 할 것인지, 인식대상 종의 수에 따라 식별이 가능할 수 있을 만큼 얼마나 정밀하게 모델의 구조가 정의 되어야 하는 지 등의 문제가 존재한다.

본 연구에서는 이러한 문제를 해결하기 위하여 모델 구조를 사전에 고정하지 않고 학습 및 인식 과정을 수행하면서 훈련 데이터에 적응하여 최적의 모델을 결정하고 이를 이용하여 최적의 소리모델을 생성하는 알고리즘을 고안하였다. 여기서 최적의 소리모델이란 HMM 모델의 상태수와 GMM의 수를 자동으로 탐색하여 전체 데이터의 인식률을 최고의 수준으로 향상시키는 방법을 의미한다.

동물소리를 이용하여 동물을 식별하는 시스템

에서 입력되는 소리는 한 종의 소리가 두드러지고 미미한 다른 소리가 포함될 수 있다는 사실에 기초하여 최대빈도 소리를 탐색하고 이것으로 소리의 종을 판단하는 방법을 고안 하였다. 즉 비터비 탐색에서 탐색범위를 최대로 하여 모델리스트를 얻고 이 리스트에서 최대빈도의 모델을 탐색하여 인식결과로 하는 최대 빈도모델 탐색알고리즘을 제안하였다.

본 논문의 구성은 다음과 같다. 2장에서 잡음 등을 제거하고 정리하는 소리 분리과정, 소리의 특징추출과정, 학습과정 및 비터비 탐색 과정으로 구성되는 동물소리 모델 개발 시스템을 소개한다. 3장에서 최대 빈도모델 탐색 알고리즘을 기술하고 4장에서 소리 모델생성 알고리즘을 기술한다. 5장에서 실험내용을 정리하고 6장에서 연구결과 및 향후 계획을 정리한다.

2. 동물소리모델 시스템 개발

본 연구에서 동물소리 인식용 소리 모델개발을 위하여 그림1과 같은 시스템을 구성하였다. 소리영역분리과정(Split)에서는 준비된 훈련용 동물소리 코퍼스(Training Sounds)와 시험용 동물소리 코퍼스(Test Sounds)에 대하여 소리영역을 분리하고, 특징추출과정(Feature Extraction)에서는 MFCC 특징벡터들을 생성한다. 학습과정(Training)에서는 추출된 MFCC 특징들을 입력으로 HMM(Hidden Markov Model)의 모델(Models) 집합을 생성하고, 비터비 탐색(Viterbi Search)과정에서는 이 모델집합과 시험용 데이터에 대하여 비터비 탐색알고리즘으로 탐색영역을 최대로 하여 최적의 모델리스트를 탐색하여 출력한다. 최대 빈도모델 탐색과정(Most Frequent Model Search)과정에서는 이 리스트에서 최대빈도의 모델을 탐색하여 최종인식 결과를 출력한다.

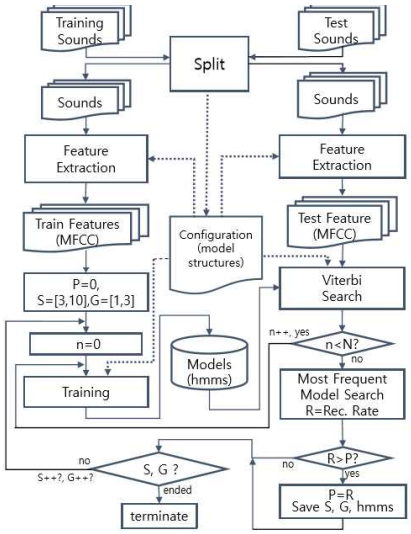

그림 1. 동물소리모델생성 시스템의 다이어그램
Fig. 1. System diagram of sound model generation

제어변수들(P,S,G)은 소리모델자동생성과정에서 이용된다. 4장에서 상세 설명된다. 제어파일(Configuration)은 특징추출, 학습, 비터비 탐색의 과정에서 요구되는 정보들을 정의해 놓은 파일들의 집합이다. 모든 과정은 이 제어파일의 수치데이터에 따라 처리가 제어된다. 주요 파일로 특징추출에 필요한 MFCC 특징의 사양, 소리데이터의 모델구조, 학습에 필요한 소리별 라벨 마스터정보 및 HMM 구조정보, 비터비 탐색에 필요한 발음사전 등 있다.

표 1. 모델생성과 실험을 위한 소리데이터
Table 1. Sound data for model generation and test.

ID	Species	Train		Test		All
		IDs	No	IDs	No	
B00	Chaffinch	100	3	216	1	4
B01	Goldcrest	103	3	217	2	5
B02	Grasshopper Warbler	106	3	219	1	4
B03	Oystercatcher	109	3	220	2	5
B04	Pheasant	112	6	222	4	10
B05	Siskin Finch	118	5	226	2	7
B06	Sparrow	123	4	228	2	6
B07	Yellow Hammer	127	8	230	4	12
B08	Baikal Teal	135	4	234	1	5
B09	Black-browed Reed Warbler	139	3	235	2	5
B10	Swan	142	4	237	2	6
B11	Crow	146	3	239	1	4
B12	Brown Dipper	149	6	240	3	9
B13	Magpie	155	5	243	2	7
B14	Bull-headed Shrike	160	6	245	3	9
B15	Azure-winged magpie	166	4	248	1	5
B16	Black-naped Oriole	170	7	249	2	9
B17	Daurian Redstart	177	4	251	2	6
B18	Great Tit	181	4	253	2	6
B19	Common cuckoo	185	5	255	2	7
B20	Eastern rowned Warbler	190	4	257	2	6
B21	Great Egret	194	4	259	2	6
B22	Dog	198	3	261	2	5
B23	Horse	199	3	263	1	5
B24	Pig	204	4	265	2	5
B25	Wolf	208	5	266	2	7
B26	Cat	213	3	268	2	5
All			116		54	170

2.1 소리분리과정(Split)

동물의 소리를 구축하기 위해서는 자연에서 소리를 녹취하게 되고 이렇게 녹취된 소리데이터에는 주변 환경의 잡음 또는 다른 종의 소리가 더해질 수 있으므로 정리과정이 필요하다. 훈련 및 시험을 위한 데이터는 오디오 편집 시스템을 이용하여 청취 및 파형을 보면서 잡음을 제거하고 학습에 적절하도록 절단할 필요가 있다. 특히 동물 소리의 경우에는 소리의 에너지를 중심으로 분리하여 모델을 학습시키는 것이 효율적이다.

소리 분리 과정에서 모든 데이터 원본을 오디오 편집시스템을 이용하여 잡음여부, 절단 등을 판단하여 학습용 및 시험용 소리 데이터를 표1과 같이 작성하였다. 총 27종의 동물 소리데이터로 학습용 데이터 116개와 시험용 데이터 54개 총 170개를 준비하였다. 소리데이터는 종별 고유번호(ID)와 데이터 일련번호(IDs)를 갖으며 학습용(Train) 및 시험용(Test)로 분류된다.

1

2.2 특징추출과정(Feature Extraction)

특징추출과정에서는 소리데이터들을 MFCC 알고리즘[12]을 이용하여 MFCC 벡터집합으로 변환한다. 소리데이터는 44100Hz 양자화비율, 스테레오 채널, 16비트 샘플의 PCM Wave 포맷이다. MFCC 벡터의 기본구조는 12차 켑스트럼(Cepstrum) 계수에 1차 에너지를 추가하여 13차 벡터이고 델타연산 및 2차미분의 연산을 추가하여 39차 벡터를 사용하였다. 이 연산에는 고주파 영역의 신호를 강조하기 위하여 필요한 값(rpe-emphasis) 0.97, 250ms의 해밍윈도우(Hamming Window)와 10ms 이동률이 사용되었다. 그림 1에서와 같이 모든 학습용 및 시험용 소리데이터 집합들이 일괄적으로 한 번에 MFCC로 변환되어 저장되고 다음 단계의 연산에서는 이 MFCC 벡터들을 소리데이터로 사용한다.

2.3 학습과정(Training)

학습과정에서는 훈련 데이터집합(116개)으로 HMM 모델(27개)집합을 학습한다. 학습과정에 소리데이터별로 소리에 해당하는 모델 열을 기술하고 있는 표2와 같이 소리데이터의 모델구조 스크립트를 사용한다. 표1에 27종의 동물 종별 고유번호(ID)가 지정되어 있으며 전체 소리데이터는 일련번호(IDs)를 추가하여 데이터고유번호가 된다. 예를 들어 되새의 종은 종 고유번호가 B00이고 3개의 소리데이터가 100, 101, 102의 일련번호를 가지고 있으므로 첫 번째 소리데이터의 데이터고유

번호는 B00_100이고 이 데이터의 모델이름은 종의 고유번호와 동일하게 B00로 정의한다. 두 번째 소리데이터의 고유번호는 B00_101이고 모델은 B00이다. 이 와 같은 방법으로 27종의 종 고유번호와 170개 소리데이터의 데이터 고유번호 및 모델이 지정된다.

표 2. 소리데이터에 대한 모델고유번호와 모델리스틀 수록한 모델구조 스크립트

Table 2. Model structure script that consists of sound model identifier and a list of models for sound data.

```
B00_100 B00 B00
B00_101 B00 B00 B00
B00_102 B00 B00 B00 B00
B01_103 B01 B01
B01_104 B01 B01 B01
...
B21_197 B21
B00_198 B00 B00 B00 B00 B00 B00
B01_199 B01 B01
...
B26_269 B26
```

27종의 종별고유번호는 B00부터 B26으로 정의되고 학습용 및 시험용에 대해서도 동일한 종별고유번호를 사용한다. 표2에서 소리데이터의 고유번호 B00_100은 B00 B00이라는 두 개의 모델로 즉 이 소리데이터가 2개의 소리 모델로 구성되어 있음을 의미 한다. 학습용 소리데이터는 B00_100부터 B26_215까지, 시험용 소리데이터는 B00_216부터 B26_269까지이고 모든 소리데이터의 구성 모델열을 정의하여야 한다. 이 스크립트는 소리분리 과정에서 오디오 편집시스템을 이용하여 소리데이터의 파형을 참고하여 작성하였다. 이 모델구조 스크립트는 학습용으로 사용되는 마스터라벨파일, 모델리스트, 문법작성의 기준이 된다.

학습과정은 표2와 같이 정의된 모델구조에 근거하여 소리데이터에 대하여 해당 HMM모델이 최대의 관측확률이 얻어지도록 HMM의 상태 천이 확률과 GMM의 값을 추정해가는 과정이다. 학습에 앞서 정의되어야하는 HMM모델의 구조는 일련의 상태들과 상태에 소속되는 GMM 및 상태천이확률로 구성된다. 상태의 수 S는 연결용으로 사용되는 2개의 상태를 포함하여 3이상의 상태를 갖게 된다. 각 상태는 G 세트의 GMM를 가지고 있다. 1 세트의 GMM(Gaussian Mixture Model)은 39차의 공유분산, 중앙값 및 가중치로 구성된다. 상태천이확률은 SxS 크기의 배열로 표현된다. 상태의 수 S와 GMM의 수 G는 소리모델생성알고리즘에 의하여 조정되고 <S,G>의 값이 지정된 상태에서 학습이 이루어진다. 학습은 Baum-Welch 알고리즘[13]을 사용하여 수행되며 학습의 원리는 다음과 같다. 예를 들어 소리모델 B00의 학습에 소리데이터 B00_100, B00_101, B00_!01가 사용된다. 표2에 근거하여 소리데이터 B00_100은 두개의 모델 B00 B00로 정의되어 있으므로 이 소리데이터로는 두 개의 모델을 훈련한다. 즉 소리데이터 B00_100의 특징 MFCC 벡터열에 대하여 관측확률이 최대가 되도록 두개 HMM의 구조의 수치(GMM의 정보, 상태의 천이확률)을 추정하여 개선한다. 소리데이터 B00_101도 B00, B00, B00에 대하여 재추정하고, 소리데이터 B00_102도 B00 B00 B00 B00에 대해서도 다시 재 추정한다. 이와 같은 방법으로 B00의 HMM학습이 완성된다.

상태의 수와 GMM의 수 <S,G>가 고정된 상태에서 27종의 모델을 116개의 훈련데이터로 학습시킨다. 이 때 요구되는 구성(configuration)파일로 훈련용 소리데이터의 목록을 수록한 훈련스크립트, 모든 소리데이터에 대한 모델을 라벨링한 마스터라벨파일(표2로부터 생성된다)이 필요하다. 예를 들어 소리데이터 B00_100에 대한 마스터라벨은 다음과 같다.

"*/B00_100*.lab"\n sil\n B00\n sil\n B00\n sil\n .\n

소리데이터 B00_100*.wav 파일은 두 개의 모델 B00 B00과 묵음 모델(sil)을 훈련하게 된다.

HTK[10]의 HeRest[11] 추정연산은 훈련스크립

트의 소리데이터 집합을 마스터라벨파일의 해당 모델 열에 따라 일괄 훈련한다. 표3은 <S,G>=<10, 1>인 경우 추정연산(학습)과정을 7회 반복하여 얻어진 모델집합으로 인식률을 계산한 것으로 반복횟수에 따라 인식률이 개선되어짐을 보이고 있다. 훈련데이터(Train)와 시험용 데이터(Test) 에 대하여 수행되었고 평균값(Avg)을 전체 인식률로 사용한다. 6회 째 학습결과의 인식률이 최대임을 보이고 수 있다. 그림2는 추이를 그래프로 표시한 것으로 인식률이 일정 수(6회)의 학습 후에는 개선이 멈추는 것을 알 수 있다.

표 3. 7회 재학습의 인식결과
Table 3. Recognition results of 7 training processes.

	1	2	3	4	5	6	7
Train	75.86	97.41	95.69	96.55	95.69	**98.28**	99.14
Test	64.81	87.04	88.89	90.74	94.44	**96.30**	92.59
Avg(R_n)	76.13	92.23	92.29	93.65	95.07	**97.29**	95.87

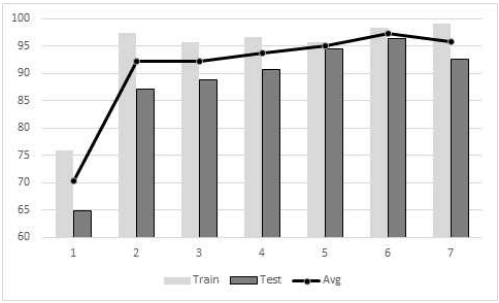

그림 2. 7회 재학습의 인식결과
Fig. 2. Recognition result of 7 training processes

2.4 비터비 탐색과정(Viterbi Search)

인식방법은 비터비 탐색알고리즘[14]을 이용한다. 비터비 탐색알고리즘은 학습된 모델집합의 모델조합의 모델로 가능한 모든 모델열의 집합 중에서 입력소리데이터에 대하여 최대 확률을 출력할 수 있는 하나의 모델 열을 탐색한다. 탐색 대상의 모델열 집합은 다음과 같은 문법으로 지정하였다.

$word = B00 | B01 | ... | B26 ;
(<$word | sil>

즉 탐색대상이 27개의 모델과 묵음모델(sil)로 가능한 모든 조합의 모델열 집합이다. 비터비연산은 HTK의 HVite 연산[11]을 이용하였다. 상기 문법과 훈련데이터를 입력으로 비터비 탐색연산을 하여 표4와 같은 형식의 탐색결과를 출력한다. 예를 들어, 소리데이터 B00_100.wav의 경우 B00 B00 B00의 3 개 모델의 열이 최적으로 탐색되었음을 보이고 있다. 묵음 모델 sil의 탐색결과에서 생략하였다.

표 4. 비터비 탐색과정의 출력결과
Table 4. Result of Viterbi Search Process

```
"../B00_100.rec"\n B00\n B00\n B00\n .
"../B00_101.rec"\n B00\n B00\n B00\n B00\n .
"../B00_100.rec"\n B00\n B00\n .
"../B00_101.rec"\n B00\n B00\n B00\n
"../B00_102.rec" B00\n .
...
...
"../B03_203.rec" B03\n B03\n B03\n B03\n B01\n
B03\n B00\n B16\n .
...
```

3. 최대 빈도 모델 탐색알고리즘

동물소리를 이용하여 동물을 식별하는 시스템에서는 입력되는 소리가 한 종의 소리가 두드러지고 소수의 다른 소리가 미미하게 포함될 수 있다는 사실에 기초하여 최대빈도 소리를 탐색하여 소리의 종을 식별하는 방법을 제안하였다.

비터비 탐색 알고리즘이 출력한 최적의 모델열의 모델 빈도수를 계산하여 최대 빈도수의 모델을 인식결과로 판단한다. 예를 들어 표4의 비터비탐색은 소리데이터 B03_203.wav에 대한 최적의 모델 열로 8개의 모델의 열 <B03 B03 B03 B03 B01 B03 B00 B16>로 판단하였다. 모델 B03의 빈도수가 5로 최대가 되므로 이 소리데이터

B00_203의 최대 빈도모델 탐색 연산의 결과는 B03이 된다. 다음은 비터비 탐색의 출력 모델열(List)로부터 최대 빈도수의 모델(model)을 계산하는 Linq 형식의 최대 빈도모델 탐색연산(Mfms)이다.

```
Function Mfms(List)
{
 var grps = (from item in List
             group item by item into g
             select new{
           Key= g.Key,
           cnt = grp.Sum(x => 1)
             }).OrderByDescending(p=>p.cnt);
   model=grps.ToList()[0].Key;
   return model;
}
```

최대 빈도모델 탐색과정에서는 HHM의 상태수와 GMM의 수 <S,G>가 고정된 상태에서 그림1과 같이 학습 및 비터비 탐색과정을 N회 반복하여 얻어지는 모델집합 H^n와 비터비 탐색결과 파일 M_0(훈련용 데이터의 비터비 탐색결과), M_1(시험용의 결과)을 이용하여 인식결과의 최대값(R)과 해당 모델집합(hmms)을 출력한다. N번 다음은 최대 빈도모델 탐색과정의 상세 내용이다. 시행된 결과의 집합을

$S=\{(M_0^n,M_1^n,H^n)|n=1..N\}$이라고 하자.

1. 입력: S, 고정 <S, G>의 비터비탐색결과
   S의 모든 (M_0^n,M_1^n,H^n)에 대하여

1.1. (M_0^n,M_1^n,H^n)의 모든 M_j^n 대하여

$Cor[1..L]=0,\ Err[1..L]=0$

1.1.1. M_j^n(표4 참조)를 읽어 들이고 각각의 데이터 L에 대하여

1.1.1.1. L에서 소리데이터 ID, 모델 M, 및 모델리스트 List를 분리한다.

1.1.1.2. $model=Mfms(List)$
$if(M==model)\ Cor[M]++$
$else\ Err[M]++$

n번 시행, j번째 데이터에 대한 인식률

$$R_{n,j}=\frac{\sum_{i=1}^{L}Cor[i]}{\sum_{i=1}^{L}Cor[i]+\sum_{i=1}^{L}Err[i]}$$

1.2. n번째 시행의 인식률 계산

$R_n=(R_{n,0}+R_{n,1})/2$

. <S,G>의 최대 빈도 모델 탐색과정의 인식률 R, 최적의 모델집합 hmms를 출력한다.

$t=\underset{n\in[1..N]}{\arg\max}\{R_n\}$

$R=R_t, hmms=H^t$

L로부터 모델리스트분리의 예로, 표4에서 L이 "../B00_100.rec"\n B00\n B00\n B00\n . 인 경우 소리데이터 ID는 B00_000 이고 모델 M(종의 고유번호)은 B00 그리고 모델리스트 List=<B00, B00, B00>이다.

표 3은 <S,G>=<10,1>인 경우 최대 빈도모델 탐색과정을 보이고 있다. 7회 수행하여 얻어진 결과로 평균값(Avg)들이 매회 계산된 인식률 R_n이고 6회 째의 결과가 최대이므로 최대 빈도모델 탐색과정의 최종 값은 $R=97.29$이 된다.

4. 소리 모델생성 알고리즘

소리 모델생성 알고리즘은 그림1과 같이 HMM의 수 S와 GMM의 수 G를 제어하면서 학습, 비터비 탐색 및 최대 빈도모델 탐색과정을 반복하여 인식률이 최대가 되는 모델집합을 탐색한다. 상세내용은 다음과 같다.

알고리즘 기술을 위하여 다음의 변수가 사용된다. 최대 확률 P, HMM의 상태수와 GMM의 세트수(S, G), 상한(St, Gt), 최대값의 인덱스(Sm,Gm).

1. 탐색범위를 초기화한다. P=0, S=nSs, St=nGt G=1, Gt=nGt, Sm=-1, Gm=-1
2. 반복변수 n=0
2.1. 학습과정(Training) 수행 및 hmms 생성
2.2. 비터비(Viterbi)탐색과정 수행, 결과

(M_0^n, M_0^n, H^n) 저장

2.3. if(n < N) 2.1로 이동

3. $\{(M_0^n, M_1^n, H^n) | n = 1..N\}$으로부터 최대 빈도모델 탐색과정수행을 수행하여 인식률 계산 R 및 hmms 계산한다.

4. if(R>P) { P=R, Sm=S, Gm=G,
   Hmms =hmms}

5. if(S < St) S++
   else {G++, S=St=Sm; }

6. if(G<=Gt) 2번 과정으로 이동한다.

7. 알고리즘이 종료되면 Hmms을 생성된 소리모델로, HMM구조는 <Sm,Gm>로 출력 한다.

5. 실험 및 결과

제안한 동물 소리모델생성알고리즘을 검증하기 위하여 다음과 같이 실험을 수행하였다. 표1과 같이 총27종의 동물소리에 대하여 훈련데이터 116개, 시험용 데이터 54개 총 170개를 준비하여 실험을 수행하였다.

제안된 소리 모델생성 알고리즘을 구현하기 위하여 특징추출과정 , 학습과정 및 비터비 탐색과정은 HTK의 연산을 사용하였으며 최대 빈도모델 탐색알고리즘 및 소리모델생성알고리즘은 C# 언어로 구현 하였다. 실험시스템의 전체구성은 그림 1과 같다.

분리과정(Split)에서 잡음을 제거하고 편집하여 표1과 같이 훈련용 데이터 116개와 시험용 데이터 54개를 준비하고, 소리데이터별 모델구조 스크립트를 표2와 같이 작성하였다.

특징추출과정에서 훈련용 및 시험용 데이터 전체를 MFCC 특징 벡터 열로 변환하여 저장한다. 소리데이터의 형식은 샘플링은 44100Hz, 스테레오 채널, 16비트 크기의 샘플이고, 소리특징 MFCC는 기본 켑스트럼 계수 12차와 1차의 에너지 그리고 이에 대한 델타연산 및 2차 미분연산을 이용하는 39차의 벡터구조이다.

학습과정에서 훈련용 데이터를 주어진 구조의 HMM모델로 훈련하여 27종의 HMM모델을 출력한다. 비터비 탐색과정에서 학습된 모델과 훈련용 데이터로부터 소리 데이터별로 모델 열을 탐색하여 표4와 같이 출력한다. 최대 빈도모델 탐색과정에서 7회의 학습에 대한 비터비 탐색결과 파일의 집합으로부터 최대인식률을 계산하여 출력한다.

탐색영역을 설정하기 위하여 Ss=4, St=13, Gt=3으로 정하고 소리모델생성알고리즘을 수행하여 학습용 및 시험용 데이터로 실험을 진행하였다. 표5는 수행과정을 도표화 한 것으로 <S,G>=<3,1>에서는 최대 빈도모델 탐색의 인식률 R=72.19 퍼센트로 시작하여 상태수가 10이 될 때까지 인식률이 점증하고 그 이후에는 정지 또는 하락하는 추세를 보였다. 알고리즘은 최대값을 P에 정확히 저장하였고 이후 GMM의 수가 증가하여도 더 이상이의 개선 없이 종료하였다.

표 5 소리모델생성알고리즘의 수행과정
Table 5. Progress of the sound model generation algorithm.

S,G	R	P	S,G	R	P
3,1	72.19	72.19	**10,1**	**97.29**	**97.29**
4,1	91.00	91.00	11,1	97.29	97.29
5,1	92.42	92.42	12,1	95.43	97.29
6,1	93.71	93.71	13,1	96.79	97.29
7,1	94.14	94.14	10,2	95.93	97.29
8,1	95.87	95.87	10,3	97.29	97.29
9,1	96.79	96.79			

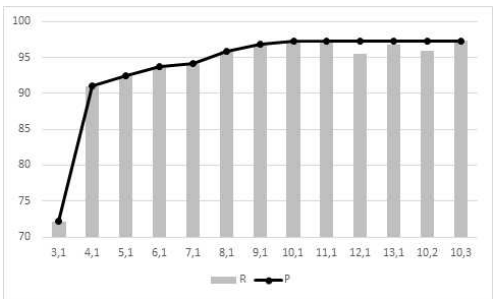

그림 3 소리모델생성알고리즘의 수행과정
Fig. 3. Progress of the sound model generation algorithm.

계산횟수는 상태 수 범위와 GMM의 수 범위의

합이다. 이 경우 (13-3)+(4-1)=13이다. 표에서와 같이 상태 수는 전 영역(3 부터 13)을 탐색하지만 GMM은 최대값의 상태 수 10은 고정하고 탐색한다.

그림3에서 막대는 현재의 최대 빈도모델과정의 인식률 R이고 실선은 최대값 P를 보이고 있다. 인식률의 개선추이가 <10,1>이후에는 정지함을 알 수 있다. 알고리즘은 <10,1>의 HMM으로 27개의 소리모델을 최대 확률 97.29 퍼센트로 계산해 냄을 확인하였다.

실험의 최종결과는 HMM의 상태 수 10, GMM의 수 1이 최적의 값으로 탐색되어 되었고, 학습용 데이터에 대하여 98.28퍼센트, 시험용 데이터의 경우 96.30퍼센트로 평균 97.29퍼센트의 인식률을 갖는 27개의 HMM 모델의 소리모델집합이 생성됨을 확인하였다.

Table 6 생성된 모델의 정보
Table 6. Information of the generated model set.

Model	State	GMM	Train	Test	Average
27	10	1	98.28	96.30	97.29

6. 결론

본 논문에서는 최대 빈도모델 탐색알고리즘을 고안하고 이를 이용한 소리모델생성 알고리즘을 제안하였다. 소리모델생성 알고리즘은 그림1과 같이 HMM의 구성요소인 상태 수 S와 GMM의 수 G의 영역을 탐색하여 최고의 인식률을 얻을 수 있는 HMM 구조를 탐색한다. 즉 학습, 비터비탐색 및 최대 빈도모델 탐색과정을 반복하며 HMM 구성요소<S,G>를 탐색하여 인식률이 최대가 되는 모델집합을 생성한다.

동물소리를 이용하여 동물을 식별하는 시스템에서는 입력되는 소리가 한 종의 소리가 두드러지고 미미한 다른 소리가 포함될 수 있다는 사실에 착안하여 최대빈도 소리를 탐색하여 소리의 종을 식별하는 방법을 고안 하였다. 비터비 탐색에서 탐색범위를 최대로 하여 모델리스트를 얻고 이 리스트에서 최대 빈도모델을 탐색하여 인식결과로 결정하는 방법이다.

알고리즘의 검증을 위하여 훈련데이터 116개 시험용 데이터 54개 총27종의 동물소리 170개를 준비하여 실험을 수행하였다. 소리모델생성알고리즘의 탐색영역을 HMM상태수를 3부터 13까지, GMM의 수를 1부터 3까지로 설정하여 수행된 실험에서 상태 수는 10, GMM수는 1로 탐색되었고, 훈련용 데이터의 경우 98.28퍼센트, 시험용 데이터의 경우 96.30퍼센트로 평균 97.29퍼센트의 인식률을 갖는 소리모델집합이 생성됨을 확인하였다.

모바일 애플리케이션으로 소리를 녹음하고 인식할 수 있는 시스템을 개발하고 인식 시도된 소리데이터를 서버에 수록하고, 이 중 품질이 검증된 데이터는 시스템의 성능개선용으로 사용하는 환경을 구축할 계획이다.

REFERENCES

[1] C. Lee, Y. Lee, Z. Ren, “Automatic Recognition of Bird Songs Using Cepstral Coefficients”, *Journal of Information Technology and Applications* Vol. 1 No. 1, May, pp.17-23, 2006

[2]. D. Mane, Rashmi R.A., S. L. Tade, “Identification & Detection System for Animals from their Vocalization”, *International Journal of Advanced Computer Research*, vol. 3. pp.352 - 357. 2013

[3] D. Mitrovic and M. Zeppelzauer, “Discrimination and retrieval of animal sounds,” *IEEE Multimedia Modelling Conference,* 2006.

[4] G. G. and Z. Li., “Content-based classification and retrieval by support vector machines,” *IEEE Transactions on Neural Networks*, vol. 14, pp. 29 - 215, 2003.

[5] H. Chen, C. Huang, Y. Chen, C. Chen, and S. Chien, “An Intelligent Nocturnal Animal

Vocalization Recognition System“, *International Journal of Computer and Communication Engineering*, Vol. 4, No. 1, pp.39 - 45, 2015

[6] Chou, C,. and Liu, P,. (2009). “Bird Species Recognition by Wavelet Transformation of a Section of Birdsong”, *Proceeding Of symposia and workshop on ubiquitous, Autonomies and Trusted Computing*. pp 189-193.

[7] Z. Le-Qing, “Insect sound recognition based on MFCC and PNN”, *2011 International Conference on Multimedia and Signal Processing*, pp. 42-46, 2011

[8] L. Rabiner and B. H. Juang. *Fundamentals of speech recognition*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.

[9] I. S. Hong, Y. J. Ko, H. S. Shin, Y. J. Kim, “Emotion Recognition from Korean Language using MFCC, HMM, and Speech Speed”, *The 12th International Conference on Multimedia Information Technology and Applications(MITA2016)*, pp.12-15, 2016

[10] Hidden Markov Model Toolkit, http://htk.eng.cam.ac.uk/. (accessed Jan., 10, 2017)

[11] S. Young, etal, “The HTK Book (for HTK Version 3.4)”,Cambridge University Engineering Department, 2009

[12] MFCC(Mel-Frequency Cepstral Coefficients) Algorithm, https://en.wikipedia.org/wiki/Mel-frequency_cepstrum, (accessed Jan., 26,2017)

[13] Baum-Welch Algorithm, https://en.wikipedia.org/wiki/Baum-Welch_algorithm, (accessed Jan., 26,2017)

[14] Viterbi Algorithm, https://en.wikipedia.org/wiki/Viterbi_algorithm, (accessed Jan., 26,2017)

저자약력

고 유 정(Ko You Jung) **[정회원]**

- 2004년 2월 : 한밭대학교 컴퓨터공학과(공학석사)
- 2009년 2월 : 한밭대학교 컴퓨터공학과(공학박사)
- 2006년 3월 ~ 현재 : 한밭대학교 시간강사

<관심분야> 음성인식, 웹 서비스

김 윤 중(Kim Yoon Joong) **[정회원]**

- 1981년 2월 : 충남대학교 전자공학과(공학석사)
- 1989년 2월 : 충남대학교 전자공학과(공학박사)
- 1984년 3월 ~ 현재 : 한밭대학교 컴퓨터공학과 교수

<관심분야> 음성인식, 인공지능, 웹 서비스, 디지털 신호처리,