Matrix Factorization을 이용한 음성 특징 파라미터 추출 및 인식

이광석* · 허강인**

Feature Parameter Extraction and Speech Recognition Using Matrix Factorization

Kwang-Seok Lee* · Kang-In Hur**

요 약

본 연구에서는 행렬 분해(Matrix Factorization)를 이용하여 음성 스펙트럼의 부분적 특징을 나타낼 수 있는 새로운 음성 파라미터를 제안한다. 제안된 파라미터는 행렬내의 모든 원소가 음수가 아니라는 조건에서 행렬분해 과정을 거치게 되고 고차원의 데이터가 효과적으로 축소되어 나타남을 알 수 있다. 차원 축소된 데이터는 입력 데이터의 부분적인 특성을 표현한다. 음성 특징 추출 과정에서 일반적으로 사용되는 멜 필터뱅크(Mel-Filter Bank)의 출력을 Non-Negative 행렬 분해(NMF: Non-Negative Matrix Factorization) 알고리즘의 입력으로 사용하고, 알고리즘을 통해 차원 축소된 데이터를 음성인식기의 입력으로 사용하여 멜 주파수 켑스트럼 계수 (MFCC: Mel Frequency Cepstral Coefficient)의 인식결과와 비교해 보았다. 인식결과를 통하여 일반적으로 음성인식기의 성능평가를 위해 사용되는 MFCC에 비하여 제안된 특징 파라미터가 인식 성능이 뛰어남을 알 수 있었다.

ABSTRACT

In this paper, we propose new speech feature parameter using the Matrix Factorization for appearance part-based features of speech spectrum. The proposed parameter represents effective dimensional reduced data from multi-dimensional feature data through matrix factorization procedure under all of the matrix elements are the non-negative constraint. Reduced feature data presents part-based features of input data. We verify about usefulness of NMF(Non-Negative Matrix Factorization) algorithm for speech feature extraction applying feature parameter that is got using NMF in Mel-scaled filter bank output. According to recognition experiment results, we confirm that proposed feature parameter is superior to MFCC(Mel-Frequency Cepstral Coefficient) in recognition performance that is used generally.

키워드

Speech Feature Extraction, Non-Negative Matrix Factorization, Mel-scaled Filter Bank

Ⅰ. 서 론

인간이 사물을 인지하는 메커니즘은 사물의 전체 형상을 인지하는 것이 아니라 의미 있는 부분적인 특징을 인지하게 된다. 이러한 인간 뇌의 인지 과정에 대한 메커니즘을 컴퓨터를 통해 구현하고자 하는 노력이 계속되고 있으며 이러한 노력 중 NMF(Non-Negative Matrix Factorization) 알고리즘은 인간의 인지과정에 기본 개념을 둔 뇌 과학(Brain Science)의 한 분야라 할 수 있다[1-7].

전체 사물의 형상은 의미 있는 부분적인 특징들에 대한 가중치 합(weighted Sum)으로 표현될 수 있고, 이는 입력 변수의 Non-Negativity 제약을 통하여 의미 있는 부분

* 진주산업대학교 전자공학과
** 동아대학교 전자공학과

접수일자 : 2006. 2. 8

적 특징을 찾기 위해 학습되어질 수 있다. 학습과정 통해 찾아진 부분적 특징은 입력 다차원 신호의 차원축소와 특징 공간(Feature Space)의 재 표현이라는 점에서 패턴 인식(Pattern Recognition)이나 패턴 분류(Pattern Classification)에서 효과적인 특징 파라미터로 사용할 수 있다[8-10].

본 연구에서는 음성신호 스펙트럼의 멜-필터 뱅크 출력에 NMF 알고리즘을 적용하여 학습된 의미 있는 부분적 특징들을 음성 인식기의 입력 파라미터로 사용하여 기존에 일반적으로 사용되고 있는 특징인 MFCC와의 인식 성능 비교 분석을 통해서 제안된 특징 파라미터의 유용성을 검증한다.

본 논문의 구성은 다음과 같다. Ⅱ장에서 NMF 알고리즘에 대한 기본 이론과 제안된 특징추출을 위한 학습과정에 대해 설명하고, Ⅲ장에서는 제안된 특징 파라미터와 MFCC간의 인식 실험 결과 및 결과에 대한 비교 분석을 다룬다. 마지막으로 Ⅳ장에서는 결론과 향후 과제에 대해 논의한다.

Ⅱ. Matrix Factorization

2.1 배경 이론

NMF는 행렬 형태 데이터의 각 원소에 Non-Negativity 제약을 사용한 행렬분해 알고리즘으로 비교사학습(Unsupervised Learning)을 통해 행렬 V를 분해하여 행렬 W와 H의 곱으로 근사한다.

$$V \approx WH \qquad (V,M,H)_{allelement} \geq 0$$

$$V_{iu} \approx (WH)_{iu} = \sum_{a=1}^{r} W_{ia} H_{au} \tag{1}$$

식(1)에서 V는 입력 데이터 행렬$(n \times m)$, W는 가중치 특성이 있는 기저행렬(Basis Matrix) $(n \times r)$, H는 V에 대해 차원 축소된 데이터 행렬 $(r \times m)$이다. 단, n은 입력 데이터의 차원, m은 입력 데이터의 조합의 개수, r은 축소할 차원을 의미하며 $(n+m)r < nm$ 을 만족하도록 선택되어진다. 그리고 모든 행렬의 원소는 반드시 음수가 아니어야 한다.

NMF 알고리즘에서의 학습은 다음 식(2)의 목적함수(Objective Function) F가 지역 최소치(Local Minimum)로 수렴할 때 까지 분해할 행렬 W와 H를 반복적으로 갱신한다.

$$F = \sum_{i=1}^{n} \sum_{u=1}^{m} [\, V_{iu} \log (WH)_{iu} - (WH)_{iu}] \tag{2}$$

W와 H에 대한 갱신 규칙(Update Rule)은 V와 WH 간의 유클리안 거리(Euclidian Distance)를 최소화하거나 Kullback-Leibler Divergence를 최소화 하는 방법을 사용한다.

<Update Rule 1>

- Minimize Euclidian Distance $\| V - WH \|$

$$H_{au} \leftarrow H_{au} \frac{(W^T V)_{au}}{(W^T WH)_{au}}$$

$$W_{ia} \leftarrow W_{ia} \frac{(VH^T)_{ia}}{(WHH^T)_{ia}} \tag{3}$$

<Update Rule 2>

- Minimize Kullback-Leibler Divergence

$$D(V \| WH)$$

$$H_{au} \leftarrow H_{au} \frac{\sum W_{ia} V_{iu} / (WH)_{iu}}{\sum W_{ka}}$$

$$W_{ia} \leftarrow W_{ia} \frac{\sum H_{au} V_{iu} / (WH)_{iu}}{\sum H_{av}} \tag{4}$$

위의 학습 규칙에 따라 반복적으로 갱신된 $\| V - WH \|$ 와 $D(V \| WH)$는 증가하지 않으며 목적함수 F는 항상 지역최소치로 수렴하게 된다. 본 논문에서는 Kullback-Leibler Divergence를 최소화 하도록 하는 갱신 규칙 2에 의해 NMF 알고리즘을 수행하였다.

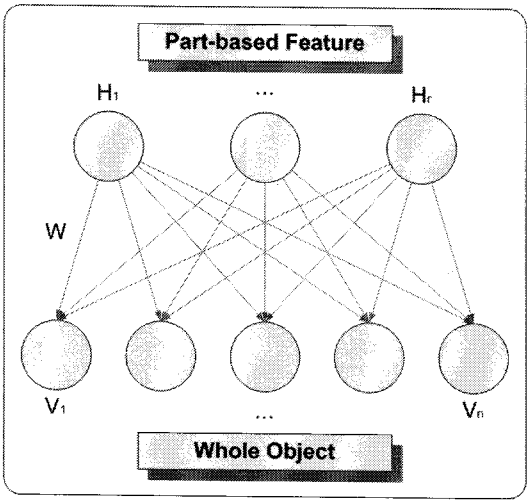


그림 1. Non-Negative Matrix Factorization
Fig. 1. Non-Negative Matrix Factorization

그림 1은 부분적 특징(Part-based Feature)들이 가중벡터 W의 선택적 가중 합을 통해 전체적인 사물의 형태를 나타내고 있는 NMF 알고리즘의 이론을 도식화 하였다. Non-Negativity를 바탕으로 학습된 가중 벡터 W는 특징 공간상에서 전체 사물의 특징에 대한 실제적인 축을 의미하며, 차원 축소된 벡터 H는 전체 사물의 특징 V에 대한 의미 있는 부분적 특징을 의미한다. 따라서 학습된 H와 W는 희소 행렬(Sparse Matrix)이 된다.

2.2 음성의 특징 요소 추출 과정

제안된 음성 특징추출 과정은 그림 2와 같다. 본 실험에서는 NMF 알고리즘의 음성신호에의 적용을 위해서 기존의 음성특징 추출과정 중 Non-Negativity를 만족하는 음성 스펙트럼의 멜 필터 뱅크 출력을 NMF의 입력으로 사용하였다. 특징 추출 과정은 다음과 같다.

해밍 창(Hamming Window)을 이용한 프레임 분석을 통해 입력으로 들어온 음성 각 프레임은 고역 강조(Pre-Emphasis)과정 후, FFT와 멜 필터 뱅크 분석을 수행하여 20차의 멜 필터 뱅크 출력을 만들어 낸다. 본 실험에서는 위의 멜 필터 뱅크 출력을 사용하여 제안된 음성 특징 벡터와 MFCC를 생성하여 인식 성능을 비교하였다.

MFCC 특징 벡터는 그림 2의 위쪽과 같이 기존의 방법과 동일하게 대수를 취한 후 DCT를 통하여 얻어진 10차의 MFCC와 얻어진 MFCC의 델타 성분 10차를 추가하여 20차의 특징벡터를 생성하였다. 그리고 제안된 음성 특징 벡터는 그림 2의 아래쪽과 같이 멜 필터뱅크 출력을 NMF 알고리즘의 입력데이터 행렬 V로 사용하였고, 학습결과 10차원으로 차원 축소된 H와 H의 Delta 성분 10차를 추가하여 MFCC와 동일한 차원인 총 20차의 특징 벡터를 생성하였다.

음성의 스펙트럼은 성대(Vocal Cord)에서 발생하는 기본 주파수(Fundamental Frequency)에 성도(Vocal Tract)의 공명현상(Resonance)에 의해 각 주파수 성분들이 추가되어 만들어진다고 가정할 때, 주파수 스펙트럼의 NMF학습 결과로 생성된 부분적 특징 H는 같은 음성에 대해 비슷한 모양을 하고 있는 성도에서 각 주파수를 발생시키는 성도 각 부분의 위치에 관한 정보를 표현하게 될 것이다. 기존에 사용하고 있는 특징 파라미터인 MFCC의 경우는

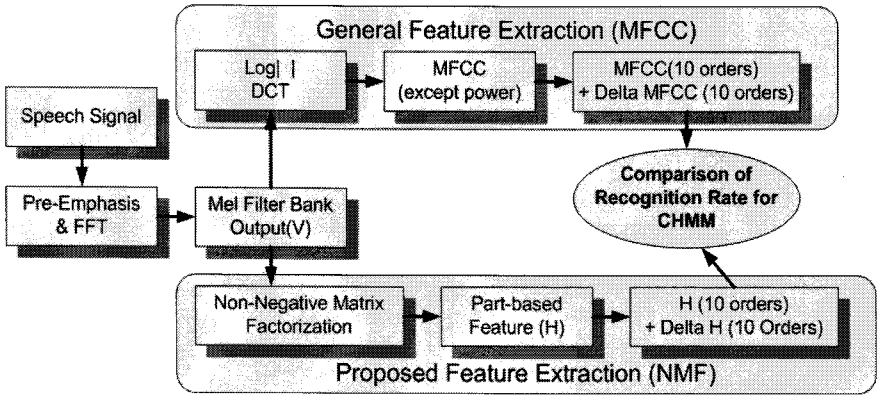


그림 2. 제안된 특징 추출 요소 과정
Fig. 2. Proposed Feature Factor Extraction Procedure

성도의 전체적인 특성을 모델링하고 있기 때문에 각각 다른 의미를 가지는 음성의 경우에도 비슷한 특성을 가지는 부분이 존재하게 된다. 그러나 NMF 학습에 의한 부분적인 특징들은 각 음성에 포함된 다양한 주파수를 발생시키는 성도 각각의 위치에 대한 모델링이기 때문에 다른 음성과의 중복성이 MFCC보다 적으며 화자간의 편차 또한 작을 것이다.

본 실험에서는 위와 같은 가정으로 음성신호 스펙트럼의 멜 필터 뱅크 출력에 NMF 알고리즘을 적용하여 적은 학습데이터에서도 음성의 부분적 특성을 통해 MFCC보다 강인한 특성을 가지는 음성 특징 파라미터를 생성하였다.

Ⅲ. 실험결과 및 분석

본 인식 실험에 사용한 음성 데이터는 ETRI Samdori 데이터 베이스로 20명의 남성 화자가 10개의 숫자음을 총 4회씩 발성한 800개의 숫자음으로 구성되어 있다.

적은 수의 학습 데이터에서 제안된 특징 파라미터의 인식성능을 확인하기 위해 학습 데이터는 10명의 화자가 각 숫자음을 1회씩 발성한 총 100개의 음성(각 숫자음 당 10개의 음성)을 사용하였고, 인식데이터는 학습 데이터를 포함한 800개의 음성 데이터 전부를 사용하여 학습 참여자(10명의 400개 음성)와 학습 미 참여자(10명의 400개 음성)로 구분하여 인식결과를 도출하였다.

특징 추출에 사용된 분석 조건은 표 1과 같다.

표 1. 분석 조건
Table. 1 Analysis Conditions

Format	16㎑, 16bit PCM
Window	Hamming(320 samples(20㎳))
Window Overlap	160 samples(10㎳)
FFT size	512 (zero padding)
Mel-Filter Bank	20
Feature Order (NMF/MFCC)	20/20 ※ included delta component

위 분석 조건에 의해 만들어진 20차의 제안된 특징 파라미터(NMF Feature)와 프레임 파워를 제외한 20차의 MFCC에 대해 인식 알고리즘으로 CHMM(Continuous Hidden Markov Model)을 사용하여 인식 실험을 수행하고 각각에 대한 인식 성능을 비교 분석하였다. 각 특징 파라미터에 대한 인식 결과는 아래의 표 2와 같다.

표 2. 인식 결과
Table. 2 Recognition Results

Feature	Recognition Data	Recognition Rate	Remark
MFCC	학습 참여자	99.25%	
	학습 미참여자	93.25%	
	Total	96.25%	
Proposal Feature	학습 참여자	99.75%	0.50% ↑
	학습 미참여자	95.25%	2.00% ↑
	Total	97.50%	1.25% ↑

실험 결과 학습 참여자 및 학습 미참여자에 대한 인식 부분 모두 제안된 특징 파라미터에 의한 인식성능이 높은 것을 볼 수 있다. 학습 참여자 인식에서는 0.50%, 특히 학습 미참여자 인식은 2.00%의 향상을 보였으며 전체적으로는 1.25%의 성능 향상을 보였다. 이의 인식 결과는 제안된 특징 파라미터가 MFCC보다 각 음성에 대한 특징의 중복성이 적고 화자 간 편차도 적었기 때문으로 생각된다.

Ⅳ. 결 론

본 연구에서는 효과적인 부분 특징 추출이 가능한 NMF 알고리즘을 사용하여 새로운 음성인식 특징 파라미터를 제안하였다. 실험 결과, 제안된 특징 파라미터에 의한 음성 인식 성능이 MFCC보다 우수함을 알 수 있었으며, 또한 적은 학습 데이터에서도 높은 인식률을 보여주었다. 이런 결과로 차후 연속 음성 인식시스템에 충분히 적용할 수 있는 가능성을 확인할 수 있었다. 그러나 특징 추출을 위한 소요시간이 기존의 특징 파라미터인 MFCC보다 길다는 점이 실시간 음성 인식 시스템에의 적용에서 다소 문제가 될 것으로 판단된다.

향후 다양한 음성 데이터베이스에 대한 다각도의 인식 실험을 통해 데이터 의존성과 파라미터 특성에 대하여 시뮬레이션을 통하여 분석하고, 또한 위의 언급된 문제점에 대한 개선 및 보완으로 실시간 연속음성 인식시스템에의 적용과 다양한 음성신호처리 분야에의 적용을 통하여 제안된 특징 파라미터에 대한 효용성을 검증해 나갈 계획으로 있다.

참고문헌

[1] Daniel D. Lee and H. Sebastian Seung, "Learning the parts of objects by non-negative matrix factorization," Nature vol. 401, Oct. 21, 1999.

[2] Daniel D. Lee, H. Sebastian Seung, "Algorithms for Non-Negative matrix Factorization," in Advances in Neural Information Procedding System 13, T. K. Leen, T. G. Dietterich, and V. Tresp, Eds., 2001.

[3] H. Y. Choi, S. J. Choi, "Learning the Sparse Codes of Speeches via Non-Negative Matrix Factorization," CVPR, 2002.

[4] Sven Behnke, "Discovering hierarchical speech features using convolutional non-negative matrix factorization," IJCNN'03, vol. 4, pp.2758-2763, 2003-10-14.

[5] Hoyer. P. O, "Non-Negative Sparse Coding," Neural Networks for Signal Processing, 2002, Proceddings of the 2002 12th IEEE Workshop on, pp. 557-565, 2002.

[6] S. Tsuge, M. Shishibori, S. Kurojwa, K. Kita, "Dimensionally Reduction Using Non-Negative Matrix Factorization for Information Retreval," Systems, Man and Cybermetics, 2001 IEEE International Conference on, vol. 2, pp. 960-965, 2001.

[7] D. Guillamet, B. Schiele, J. Vitria, "Analyzing non-negative matrix factorization for image classification," Pattern Recognition, 2002, Proceedings, 16th International Conference on, vol. 2, pp. 116-119, 11-15 Aug. 2002.

[8] L. R. Rabiner, R. W. Schafer, "Digital Processing of Speech Signals,"Prentice Hall, 1993.

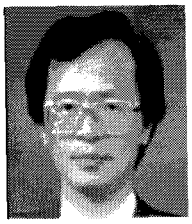
[9] L. R. Rabiner, B. H. Juang, "Fundamentals of Speech Recognition," Prentice hall, 1999.

[10] Simon Haykim, "Neural Networks a Comprehensive Foundation," Prentice Hall, 1999.

저자소개

이 광 석(Kwang-Seok Lee)

1983년 2월 동아대학교 전자공학과 (공학사)
1985년 2월 동아대학교 전자공학과 (공학석사)
1992년 2월 동아대학교 전자공학과(공학박사)
2004년 2월-2005년 1월 미국 애리조나 주립대학 객원교수
1995년-현재 진주산업대학교 전자공학과 부교수
※관심분야 : 음성신호처리 및 인식, 생체 신호처리, 지능화 기술



허 강 인(Kang-In Hur)

1980년 2월 동아대학교 전자공학과 (공학사)
1982년 2월 동아대학교 전자공학과 (공학석사)
1990년 8월 경희대학교 전자공학과(공학박사)
1984년-현재 동아대학교 전자공학과 교수
1988년 9월-1989년 8월 : 일본 객원연구원
1992년 9월-1993년 8월 : 일본 객원연구원
※관심분야 : DSP, 음성인식, 합성, 신경회로망