Non-Negative Matrix Factorization을 이용한 음성 스펙트럼의 부분 특징 추출

박정원, 김창근, 허강인
동아대학교 전자공학과

Parts-based Feature Extraction of Speech Spectrum Using Non-Negative Matrix Factorization

Jeongwon Park, Changkeun Kim, Kangin Hur
Dept. of Electronic Engineering, Dong-A University
E-mail : jwpark@donga.ac.kr

Abstract

In this paper, we propose new speech feature parameter using NMF(Non-Negative Matrix Factorization). NMF can represent multi-dimensional data based on effective dimensional reduction through matrix factorization under the non-negativity constraint, and reduced data present parts-based features of input data. In this paper, we verify about usefulness of NMF algorithm for speech feature extraction applying feature parameter that is got using NMF in Mel-scaled filter bank output.

According to recognition experiment result, we could confirm that proposal feature parameter is superior in recognition performance than MFCC(mel frequency cepstral coefficient) that is used generally.

I. 서론

최근 뇌 과학(Brain Science)분야의 연구가 활발히 진행됨에 따라 인간 뇌의 인지 과정에 대한 메커니즘이 컴퓨터에 의해 구현되고 있다. 이 중 NMF(Non-Negative Matrix Factorization) 알고리즘은 인간이 사물의 의미 있는 부분적 특징(Parts-based Feature)에 의해 전체 사물을 판단하는 인간 뇌의 인지 과정에 기본 개념을 두고 개발되었다[1].

전체 사물의 형상은 의미 있는 부분적인 특징의 가중합(weighted sum) 만으로 표현되고, 이것은 사용 변수의 Non-Negativity 제약을 사용하여 의미 있는 부분적 특징을 찾기 위해 학습시킬 수 있다.[1] 학습과정 후에 찾아진 부분적 특징은 입력 다차원 신호의 차원축소와 특징 공간(Feature Space)의 재표현이라는 점에서 패턴 인식(Pattern Recognition)이나 패턴 분류(Pattern Classification) 효과적인 특징 파라메터로 사용할 수 있다.

본 논문에서는 음성신호 스펙트럼의 멜 필터 뱅크 출력(Mel Filter Bank Output)에 NMF 알고리즘을 적용하여 학습된 의미 있는 부분적 특징들을 음성 인식기의 입력 파라메터로 사용하여 기존에 일반적으로 사용되고 있는 특징인 MFCC(Mel Frequency Cepstral Coefficient)와의 인식 성능 비교 분석을 통해서 제안된 특징 파라메터의 유용성을 검증한다.

본 논문의 구성은 다음과 같다. 2장에서 NMF 알고리즘에 대한 기본 이론과 제안된 특징추출을 위한 학습 과정에 대해 설명하고, 3장에서는 제안된 특징 파라메터와 MFCC간의 인식 실험 결과 및 결과에 대한 비교 분석을 다룬다. 마지막으로 4장에서는 결론 및 향후 과제에 대해 논의한다.

II. Non-Negative Matrix Factorization

2.1 기본 이론

NMF는 행렬 형태 데이터의 각 원소에 Non-Negativity 제약을 사용한 행렬분해(Matrix Factorization) 알고리즘으로 비교사 학습(Unsupervised Learning)을 통해 행렬(V)를 분해하여 행렬 W와 H의 곱으로 근사화 한다

$$V \approx WH \quad (V,W,H)_{all\ element} \geq 0$$

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

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

NMF 알고리즘에서의 학습은 다음 식(2)의 목적함수(Objective Function) F가 지역 최소값(Local Minimum)으로 수렴 할 때 까지 분해 할 행렬 W와 H를 반복적으로 갱신한다.

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

W와 H에 대한 갱신 룰은 V와 WH 간의 Euclidian Distance를 최소화 하거나 Kullback-Leibler Divergence를 최소화 하는 방법을 사용한다.

- **Update Rule 1**

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

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

$$W_{ia} \leftarrow W_{ia} \frac{(VH^T)_{ia}}{(WHH^T)_{ia}} \qquad (4)$$

- **Update Rule 2**

- Minimize Kullback-Leibler Divergence $D(V \| WH)$

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

$$W_{ia} \leftarrow W_{ia} \frac{\sum_u H_{au} V_{iu} / (WH)_{iu}}{\sum_v H_{av}} \qquad (5)$$

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

그림 1은 부분적 특징(part-based feature)들이 가중벡터(W)의 선택적 가중 합을 통해 전체적인 사물(whole object)의 형태를 나타내고 있는 NMF 알고리즘의 이론을 도식화하였다. Non-Negativity를 바탕으로 학습된 가중벡터(W)는 특징 공간상에서 전체 사물의 특징에 대한 실제적인 축을 의미하며, 차원 축소된 벡터(H)는 전체 사물의 특징(V)에 대한 의미 있는 부분적 특징을 의미한다. 따라서, 학습된 H와 W는 희소 행렬(sparse matrix)이 된다.

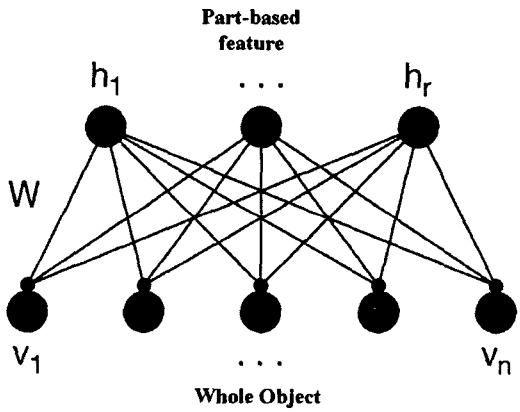

그림 1. Non-Negative Matrix Factorization

2.2 제안된 음성 특징 추출 과정

제안된 음성 특징추출 과정은 그림 2와 같다. 본 실험

에서는 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 우측과 같이 기존의 방법과 동일하게 대수를 취한 후 DCT를 하여 얻어진 10차의 MFCC에 Delta 성분 10차를 추가하여 20차의 특징 벡터를 생성하였다. 그리고 제안된 음성 특징 벡터는 그림 2의 좌측과 같이 멜 필터뱅크 출력을 NMF 알고리즘의 입력데이터 행렬(V)으로 사용하였고, 학습결과 10차원으로 차원 축소된 H와 H의 Delta 성분 10차를 추가하여 MFCC와 동일한 차원인 총 20차의 특징 벡터를 생성하였다.

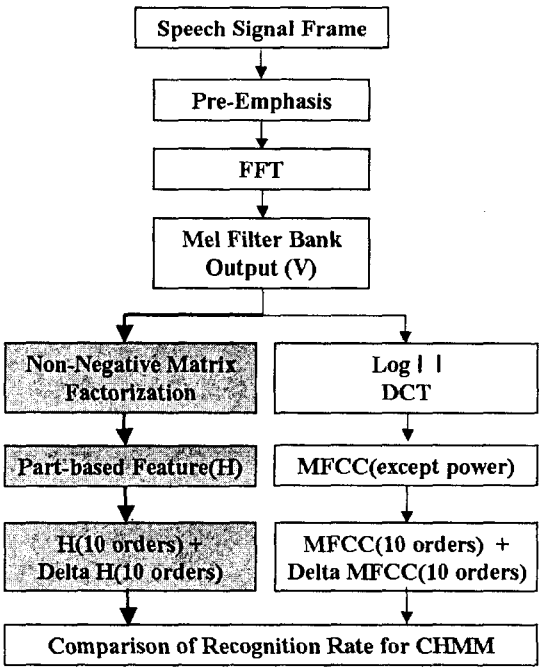

그림 2. NMF를 이용한 특징 추출 구성도

음성의 스펙트럼은 성대(Vocal Cord)에서 발생하는 기본주파수(Fundamental Frequency)에 성도(Vocal Tract)의 공명현상(Resonance)에 의해 각 주파수 성분들이 추가 되어 만들어진다고 가정할 때, 주파수 스펙트럼의 NMF 학습 결과로 생성된 부분적 특징(H)은 같은 음성에 대해 비슷한 모양을 하고 있는 성도에서 각 주파수를 발생시키는 성도 각 부분의 위치에 관한 정보라고 사료된다. 기존에 사용하고 있는 특징 파라메터인 MFCC의 경우는 성도의 전체적인 특성을 모델링하고 있기 때문에 각각 다른 의미를 가지는 음성의 경우에도 비슷한 특성을 가지는 부분이 존재 하게 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 적을 것으로 사료된다.

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 파라메터를 생성하였다.

III. 실험결과 및 비교분석

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

적은 학습데이터에서의 제안된 특징 파라메터의 인식 성능을 확인하기 위해 학습데이터는 10명의 화자가 각 숫자음을 1회씩 발성한 총 100개의 음성(각 숫자음 당 10개의 음성)을 사용하였고, 인식데이터는 학습데이터를 포함한 800개의 음성 데이터 전부를 사용하여 학습 참여자(10명의 400개 음성)와 학습 미참여자(10명의 400개 음성)로 구분하여 인식결과를 도출하였다.

특징 추출에 사용된 분석 조건은 표 1과 같다.

표 1. 분석 조건

PCM	16kHz, 16bit
Window	Hamming (320 samples,20ms)
Frame Overlap Size	160 samples (10ms)
FFT Size	512 (zero padding)
Mel-Filter Bank Num.	20
Feature Order (NMF / MFCC)	20 / 20 (include Delta Component)

위 분석 조건에 의해 만들어진 20차의 제안된 특징 파라메터(NMF Feature)와 프레임 파워를 제외한 20차의 MFCC(Mel Frequency Cepstral Coefficient)에 대해 인식 알고리즘으로 CHMM(Continuous Hidden Markov Model)을 사용하여 인식 실험을 수행하고 각각에 대한 인식성능을 비교 분석하였다. 각 특징 파라메터에 대한 인식 결과는 아래의 표 2와 같다

표 2. 인식 결과

Feature	Recognition data	Recognition Rate
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 미참여자 인식부분 모두 제안된 특징 파라메터에 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 성능향상을 보였다. 이의 인식 결과는 제안된 특징 파라메터가 MFCC보다 각 음성에 대한 특징의 중복성이 적고 화자간 편차도 적었기 때문으로 사료된다.

IV. 결론 및 향후 과제

본 논문에서는 효과적인 부분 특징 추출이 가능한 NMF 알고리즘을 사용하여 새로운 음성인식 특징 파라메터를 제안하였다. 실험 결과 제안된 특징 파라메터에 의한 음성 인식성능이 MFCC 보다 우수함을 볼 수 있었으며, 또한 적은 학습데이터에서도 높은 인식률을 보여주었다. 이런 결과로 차후 연속 음성 인식시스템에 충분히 적용가능 하다는 점을 확인할 수 있었다. 그러나, 특징 추출을 위한 소요시간이 기존의 특징 파라메터인 MFCC 보다 크다라는 점에서 실시간 음성 인식시스템에의 적용에는 아직 문제점이 남아 있다.

향후 다양한 음성 데이터 베이스에 대한 다각도의 인식실험을 통해 데이터 의존성과 파라메터 특성에 대하여 알아보며, 또한 위의 언급된 문제점에 대한 수정과 보완으로 실시간 연속음성 인식시스템에의 적용과 다양한 음성신호처리 분야에의 적용을 통하여 제안된 특징 파라메터에 대한 효용성을 검증 해보아야 할 것이다.

참고 문헌

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 Processing 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. Proceedings 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 Retrieval", 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, 1978.
9. L. R. Rabiner, B. H. Juang, "Fundamentals of Speech Recognition", Prentice Hall, 1993.
10. Simon Haykin, "Neural Networks a Comprehensive Foundation", Prentice Hall, 1999.
11. 박정원, 김평환, 김창근, 허강인, "음성 인식기 구현을 위한 SVM과 독립성분분석 기법의 적용", 대한 전자 공학회 하계종합학술대회, 26권 1호, pp. 2164-2167, 2003.