음원 희소성 추정 및 비음수 행렬 인수분해 기반 신호분리 기법

홍세린 남시연 윤덕규 최승호
서울과학기술대학교
shchoi@snut.ac.kr

A Signal Separation Method Based on Sparsity Estimation of Source Signals and Non-negative Matrix Factorization

Serin Hong Siyeon Nam Deokgyu Yun Seung Ho Choi
Seoul National University of Science and Technology

요약

비음수 행렬 인수분해(Non-negative Matrix Factorization, NMF)의 신호분리 성능을 개선하기 위해 희소조건을 인가한 방법이 희소 비음수 행렬 인수분해 알고리즘(Sparse NMF, SNMF)이다. 기존의 SNMF 알고리즘은 개별 음원의 희소성을 고려하지 않고 임의로 결정한 희소 조건을 사용한다. 본 논문에서는 음원의 특성에 따른 희소성을 추정하고 이를 SNMF 학습 알고리즘에 적용하는 새로운 신호분리 기법을 제안한다. 혼합 신호에서의 잡음제거 실험을 통해, 제안한 방법이 기존의 NMF와 SNMF에 비해 성능이 더 우수함을 보였다.

1. 서론

최근 음성인식, 음성향상, 음악검색 등의 분야에서 신호분리 기술을 이용한 잡음제거 기술의 필요성이 점증되고 있다. 신호분리란 혼합된 신호로부터 각각의 신호를 분리하는 것으로서, 최근 들어, 비음수 행렬 인수분해(Non-negative Matrix Factorization, NMF) 기법이 많이 연구되고 있다. 음성신호처리 분야에서의 NMF는 잡음이 섞인 신호의 스펙트로그램을 기저행렬(basis matrix)과 활성행렬(activation matrix)로 분리하는 것이다 [1]. 최근 들어, 신호의 희소성(sparseness)을 고려하여 기존의 NMF의 성능을 개선하기 위한 희소 비음수 행렬 인수분해 (Sparse NMF, SNMF)가 제안되었다 [2]. 하지만 기존의 SNMF는 분리하고자 하는 각 음원들의 희소성을 반영하지 않고 희소 조건을 임의로 결정하며, SNMF를 통해 최적화된 기저행렬과 활성행렬을 구하는 과정에 한계가 있다. 본 논문에서 제안하는 SNMF는 음원들의 특성에 따라 음성과 잡음의 희소성을 다르게 적용하며, 희소성은 음원의 활성행렬에 의해 결정된다.

2. 기존 NMF 및 SNMF

NMF는 비음수 행렬 V를 기저행렬과 활성행렬의 곱으로 분해하는 방식이다. 기존 NMF 알고리즘은 W와 H가 기저행렬과 활성행렬일 때, 아래의 식을 목표로 한다.

$$V \approx WH \tag{1}$$

식 (1)을 만족시키는 W와 H를 최적화하기 위해서 $D(V \| WH)$는 일반적으로 Kullback-Leibler divergence (KL 발산) 거리 함수를 통해 구한다 [3].

$$\underset{W,H \geq 0}{\operatorname{argmin}} \; D(V \| WH) + \mu \| H \| \tag{2}$$

W와 H는 증배갱신법(multiplicative update rule) 즉, 이전 갱신값에 어떤 식이 곱해지는 형태로 갱신되며[4], 식 (2)를 최소화하기 위하여 사용한다. 그리고 아래 식 (3)을 이용하여 무작위로 초기화된 행렬을 갱신하면서 최적화된 W와 H를 얻는다.

$$H \leftarrow H \frac{W^T(V/WH)}{W^T 1 + \mu}, W \leftarrow W \frac{V/WH + \widetilde{W}\widetilde{W}^T}{1 + \widetilde{W}\widetilde{W}^T(V/WH)} \tag{3}$$

여기에서 $\widetilde{W}$은 정규화된 W이다. 이 때 기존 SNMF의 경우에는 희소조건 μ를 임의로 정해서 인가한다.

3. 제안한 SNMF 방법

제안한 SNMF는 <그림 1>에서와 같이 먼저 NMF 알고리즘을 이용하여 음성신호와 잡음신호 각각에 대해 초기행렬 W_s^0, H_s^0, W_n^0, H_n^0을 구한다. 그리고 초기 활성행렬 H_s^0, H_n^0로부터 각각의 희소성 λ_s, λ_n을 구한다. 희소성 λ는 아래 식 (4)로 계산하며, N은 활성행렬 H의 원소의 개수이다.

$$\lambda = \frac{\sqrt{N} - \sum_i \sum_j |h_{i,j}| \; / \sqrt{\sum_i \sum_j h_{i,j}^2}}{\sqrt{N} - 1} \tag{4}$$

그리고 각각의 희소성 λ_s, λ_n으로 $\vec{\lambda} = [\lambda_s \lambda_s \cdots \lambda_s \lambda_n \lambda_n \cdots \lambda_n]^T$를 만든다. 본 연구에서는 각 신호별 희소성 $\vec{\lambda}$을 아래 식 (5)와 같이 식

(4)를 수정하여 적용한다.

$$H \leftarrow H \frac{W^T(V/WH)}{(W^T 1)(1+\vec{\lambda})}, W \leftarrow W \frac{V/WH + \widetilde{W}\widetilde{W}^T}{1+\widetilde{W}\widetilde{W}^T(V/WH)} \quad (5)$$

마지막으로 희소성 $\vec{\lambda}$와 초기행렬 W_n^0, H_n^0을 이용한 SNMF를 통해, 혼합 신호로부터 최적화된 W와 H를 구하고, 이를 통해 잡음이 제거된 신호를 얻는다.

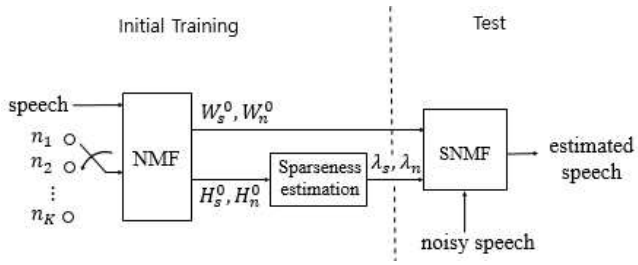

그림 1. 제안한 SNMF의 전체 흐름도

4. 실험 및 결과

기존의 NMF, SNMF 방법과 신호의 특성에 따라 희소조건을 구분하여 인가하여 제안한 SNMF 방법의 잡음 제거 성능을 비교하기 위하여 TIMIT DB [5]의 1500개 음성과 4가지 잡음 (백색잡음, 사이렌소음, 차량 주행소음, 공장소음)을 훈련하였다. 훈련에 사용한 DB와 다른 100개 음성과 잡음을 이용하여 혼합 신호를 만들어 잡음 제거 성능 실험을 진행하였다. 기존의 SNMF에서는 최적의 희소성 값을 찾기 위하여 μ = {1, 5, 10, 15} 범위에서 값을 바꿔가며 진행하였고, 평균적으로 최대가 되는 희소성을 선택하였다. 성능비교는 객관적 음질평가 도구인 PESQ [6]를 사용하였다. <표 1>의 실험 결과와 같이, 4가지 환경에서 PESQ 평균값이 기존 방법들(NMF, SNMF)보다 제안한 SNMF 방법에서 높은 점수를 보였으며, 잡음 제거 성능이 향상된 것을 알 수 있다.

표 1. 잡음 제거 성능 비교 실험 결과 (PESQ)

방법 / 잡음	처리 전	NMF	SNMF	제안한 SNMF
white	1.50	1.69	1.76	2.05
siren	1.99	2.10	2.02	2.27
car	3.15	3.29	3.44	3.40
factory	1.65	1.68	1.81	1.88
average	2.07	2.19	2.26	2.40

5. 결론 및 향후 연구방향

본 논문에서는 음원의 특성에 따라 추정한 희소성을 학습 알고리즘을 적용하는 새로운 SNMF 기반 신호 분리 기법을 제안하였다. 실험 결과, 기존의 신호 분리 기법에 비해 제안한 SNMF의 잡음 제거 성능이 우수하였다. 향후 신호분리 전처리 과정으로 딥러닝을 통한 잡음 환경을 판별하는 상황인지(context-aware) 과정을 추가하고, 목표 신호인 음성신호와 특성이 유사한 다화자잡음(babble noise) 환경에서의 SNMF 성능을 개선하기 위한 연구를 진행할 예정이다.

감사의 글

이 논문은 2017년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임. [2016-0-00144, 시청자 이동형 자유시점 360VR 실감미디어 제공을 위한 시스템 설계 및 기반기술 연구]

참고문헌

[1] C. Joder, F. Weninger, and D. Schuller, “A comparative study on sparsity penalties for NMF-based speech separation: Beyond LP-norms,” *IEEE International Conference. Acoustics, Speech and Signal Processing (ICASSP),* pp. 858-862, 2013.

[2] J. Le Roux, F. Weninger, and J. R. Hershey, “Sparse NMF: half-baked or well done?,” *Mitsubishi Electric Research Labs (MERL), Cambridge, MA, USA, Tech. Rep., no. TR2015-023,* 2015.

[3] M. N. Schmidt, and R. K. Olsson, “Single-channel speech separation using sparse non-negative matrix factorization,” *International Conference on Spoken Language Processing,* 2006.

[4] P. Smaragdis, C. Fevotte, G. Mysore, N. Mohammadiha, and M. Hoffman, “Static and dynamic source separation using nonnegative factorizations: A unified view,” *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 66-75, 2014.

[5] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, and D. S. Pallett, “DARPA TIMIT acoustic phonetic continuous speech corpus CD-ROM,” *NTIS*, 1993.

[6] RECOMMENDATION, ITU-T. “Perceptual evaluation of speech quality (PESQ): An objective method end-to-end speech quality assessment of narrow-band telephone networks and speech codecs,” *Rec. ITU-T p.862,* 2001.