

가우시안 프로세스를 이용한 음원분리

박선호^o, 최승진

포항공과대학교 컴퓨터공학과
{titan, seungjin}@postech.ac.kr

Gaussian Processes for Source Separation: WW Pseudo-likelihood Maximization

Sunho Park, Seungjin Choi
Department of Computer Science, POSTECH

음원분리 (source separation)는 신호처리분야에서 널리 연구되고 있는 근본적인 문제 중 하나이다. 음원분리는 여러 음원이 선형적으로 섞여서 관측되었을 때, 원 음원을 복원하는 기술을 의미한다. 기본적으로 독립성분분석 (ICA) 과 같은 통계적 접근방법들이 음원분리에 사용되고 있다. 그러나 이러한 방법들 중 대부분은 음원이 가지는 시간적 구조 (temporal structure)를 고려하지 않은 채, 주어진 음원이 시간에 대해서도 독립적이라는 추가적 제한 조건을 요구한다. 그러나 자연계에서 관측되는 거의 모든 신호는 시간적 구조를 가지는 시계열 (time series)데이터이므로 이러한 제한 조건은 타당하지 않다. 본 연구의 주요 목적은 위와 같이 시간적 구조를 가지는 시계열 음원이 섞여서 관측되었을 때, 효과적으로 분리해내는 음원분리 알고리즘을 목표로 한다.

음원의 시간적 구조를 고려하기 위하여 많은 방법들이 시간적 구조를 효과적으로 기술 할 수 있는 모델을 세우고 주어진 관측데이터를 통하여 이 모델을 학습하는 방법을 취한다. 대부분의 경우 모수형태 (parametric form)의 시간적 모델을 사용한다. 대표적으로 각 음원의 시간적 특성을 선형자기회귀모델 (linear auto-regression model)로 모델링하고 해당 계수들을 알고리즘 동작 중에 학습하는 방법들이 있다. 그러나 이러한 접근방법은 크게 두 가지 문제점을 지니고 있다. 첫째, 선형자기회귀모델은 특정 시계열 데이터가 가지는 비선형성을 표현하지 못한다. 둘째, 선형자기회귀모델의 계수는 실제로 음원분리와는 직접적으로 관계가 없는 nuisance parameters로 필히 marginalization이 필요하다. 그러나 기존의 선형자기회귀모델기반의 방법에서는 이러한 marginalization이 매우 어렵다. 그러므로 본 연구에서는 보다 일반적인 모델을 통하여 음원의 시간적 구조를 표현하고, 확률적 방법에 기반 하여 시간적 모델의 marginalization을 통해 오직 음원분리에만 연관된 모수를 학습하는 방법을 제안한다.

본 연구에서 각 음원의 시간적 구조는 일반적인 비선형 함수로 구현된다. 즉 현재 음원의 신호는 과거 p 개의 음원신호를 입력으로 하는 비선형 함수에 결정된다는 매우 유연한 시간 모델을 도입한다. 그리고 이 비선형 함수를 기술하기 위하여 최근 기계학습 분야에서 non-linear regression 문제에 많이 적용되고 있는 가우시안 프로세스를 이용한다. 가우시안 프로세스는 대표적인 비모수 방법으로 확률적 해석이 가능하다는 장점을 지니고 있다. 즉 확률적 해석이 가능한 kernel machine으로 이해 할 수 있다. 이는 곧 음원의 시간적 구조를 표현하는 비선형 함수에 가우시안 프로세스의 선형확률함수 (prior probability density function)를 적용함으로써 비선형 함수의 확률적 해석이 가능하다는 것을 의미한다. 특히 가우시안 프로세스의 선형확률함수는 mean function과 covariance function을 가지는 가우시안 분포 (Gaussian distribution)가 되는데, kernel function을 covariance function으로 정의함으로써 기존의 kernel machine이 가지는 여러 장점을 공유한다. 본 연구에서는 Gaussian kernel을 사용하였다.

앞서 언급한 대로, 가우시안 분포를 비선형 함수의 선형확률함수로 정의함으로써 우리는 선형자기회

귀모델이 가지는 문제점, 즉 marginalization이 어렵다는 점을 쉽게 해결할 수 있다. 결론적으로 비선형 함수는 marinalized out 되어 오직 mixing matrix (원래의 음성신호를 혼합해서 관측신호를 만드는 선형 시스템, 이를 예측할 수 있다면 원 신호의 복구도 가능하다)에 의존하는 목적식을 얻을 수 있다. 또한 이 목적식은 gradient 방법에 의해 최적화가 가능하다.

본 방법이 가지는 장점은 다음과 같다. 먼저 모델의 유연성이다. 본 방법에서 미리 사용자가 지정해 주어야 하는 모수는 비선형 함수가 얼마나 많은 과거 값을 입력으로 받는 지에 대한 모수와 Gaussian kernel의 kernel width parameter이다. 본 논문의 실험 1에서, 본 방법이 이 두 요소에 민감하지 않다는 것을 실험적으로 확인 할 수 있다. 또한 음원의 시간적 구조를 비선형 함수로 정의하였기 때문에 복잡한 비선형성을 가지는 시계열 데이터의 음원분리가 가능하다. 이는 본 논문에 수록된 실험 2에서 확인할 수 있다. 특히 복잡한 카오스 (chaotic time series)의 분리도 거의 완벽하게 가능하다.

마지막으로 본 방법의 한계점과 개선해야 될 사항은 다음과 같다. 첫째로 계산의 복잡성이다. 이는 비모수방법이 본질적으로 가지는 한계이지만, 음원분리에서는 속도가 중요한 요소이므로 이에 대한 적절한 접근이 필요하다. 그러므로 앞으로 근사적 방법을 통하여 알고리즘의 속도를 향상시키고자 한다. 그리고 kernel 기반 방법은 그 특성상 선형 지식을 학습과정에 사용하는 것이 용이하다. 음원분리에서도 특정 음원에 대한 선형지식을 통해 성능을 높이거나 특정 음원만을 추출하는 것이 가능 할 것이다. 그러므로 앞으로 선형지식을 음원분리에 적용시키는 방안을 연구하고자 한다.