강화학습 기반의 음성향상기법

박태준, 장준혁*

*한양대학교 전자컴퓨터통신공학과

e-mail: jchang@hanyang.ac.kr

Speech enhancement based on reinforcement learning

Tae-Jun Park, Joon-Hyuk Chang*

*Department of Electronic Engineering, Hanyang University

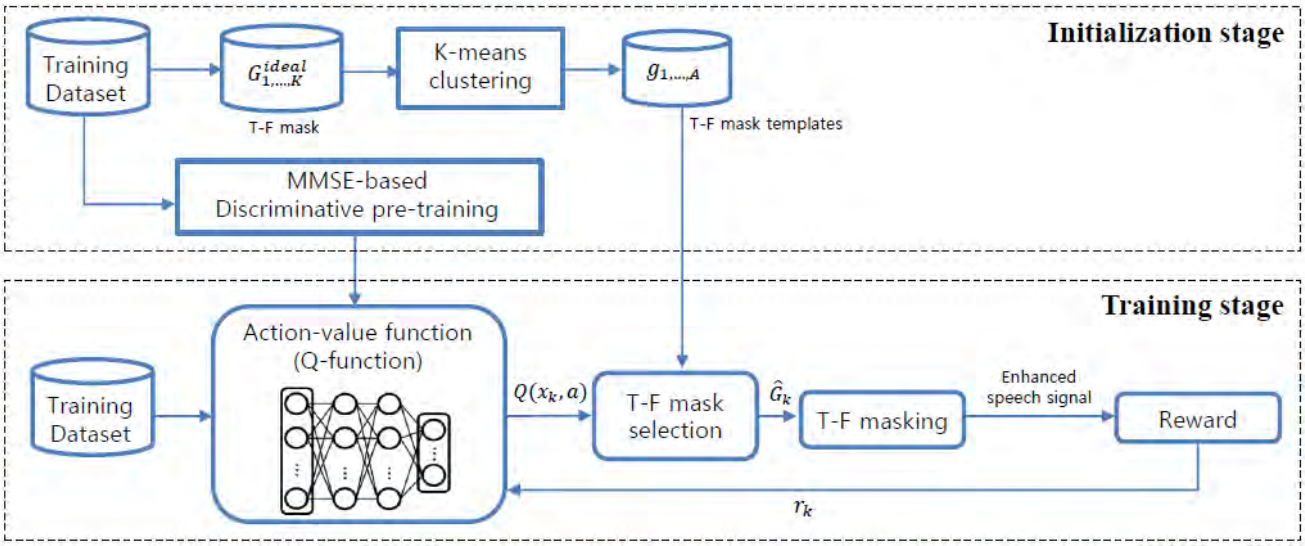
요 약

음성향상기법은 음성에 포함된 잡음이나 잔향을 제거하는 기술로써 마이크로폰으로 입력된 음성신호는 잡음이나 잔향에 의해 왜곡되어지므로 음성인식, 음성통신 등의 음성신호처리 기술의 핵심 기술이다. 이전에는 음성신호와 잡음신호 사이의 통계적 정보를 이용하는 통계모델 기반의 음성향상기법이 주로 사용되었으나 통계 모델 기반의 음성향상기술은 정상 잡음 환경과는 달리 비정상 잡음 환경에서 성능이 크게 저하되는 문제점을 가지고 있었다. 최근 머신러닝 기법인 심화신경망 (DNN, deep neural network)이 도입되어 음성 향상 기법에서 우수한 성능을 내고 있다. 심화신경망을 이용한 음성향상 기법은 다수의 은닉 층과 은닉 노드들을 통하여 잡음이 존재하는 음성 신호와 잡음이 존재하지 않는 깨끗한 음성 신호 사이의 비선형적인 관계를 잘 모델링하였다. 이러한 심화신경망 기반의 음성향상기법을 향상 시킬 수 있는 방법 중 하나인 강화학습을 적용하여 기존 심화신경망 대비 성능을 향상시켰다. 강화학습이란 대표적으로 구글의 알파고에 적용된 기술로써 특정 state에서 최고의 reward를 받기 위해 어떠한 policy를 통한 action을 취해서 다음 state로 나아갈지를 매우 많은 경우에 대해 학습을 통해 최적의 action을 선택할 수 있도록 학습하는 방법을 말한다. 본 논문에서는 composite measure를 기반으로 reward를 설계하여 기존 PESQ (Perceptual Evaluation of Speech Quality) 기반의 reward를 설계한 기술 대비 음성인식 성능을 높였다.

1. 서론

최근 인간의 여러 생활에서 매우 중요한 역할을 하고 있는 음성통신에서 주변 잡음 및 반향을 제거하는 것은 통화품질을 향상시키는데 있어서 매우 중요한 핵심적인 기술이다. 주변 잡음의 제거의 중요성이 더욱 중요시되고 이와 같은 필요성을 바탕으로 많은 잡음제거기술들이 제안되었다.

음성 향상 기법은 마이크로폰으로 입력된 잡음이 존재하는 음성 신호의 잡음을 제거하여 깨끗한 음성을 추정하는 기법으로 음성 인식과 음성 통신과 같은 음성 어플리케이션에 필수적인 기술이다. 예를 들어 음성 인식에서 잡음이 존재하지 않은 깨끗한 신호로 음성 인식 모델을 학습시킨 후 잡음이 존재하는 신호로 테스트를 할 경우 성능이 감소한다. 이를 해결하기 위하여 잡음이 존재하는 음성을 통하여 음성 인식 모델을 학습하는 방법이 제안되었으나 학습된 잡음 환경에 최적화 되어 학습된 잡음 환경에서 테스트 할 경우 우수한 성능을 보이나 학습되지 않은 잡음 환경에서 테스트 할 경우 성능이 좋지 않았다. 따라서 음성 인식 모델을 학습하기 전 잡음을 제거하기 위해 음성 향상 기법을 도입한다. 또한, 음성 향상 기법은 배경 잡음을 제거 하여 음질을 향상시키기 위하여 음성통신에서 도입되기도 하며 보청기 등에서 잡음을 제거하여 더 명확한 음성을 전달하기 위하여 도입된다. 이전에는 음성신호와 잡음신호 사이의 통계적 정보를 이용하는 통계모델 기반의 음성향상기법이 주로 사용되었으나 통계 모델 기반의 음성향상기술은 정상 잡음 환경과는 달리 비정상 잡음 환경에서 성능이 크게 저하되는 문제점을 가지고 있었다. 최근 머신러닝 기법인 심화신경망 (DNN, deep neural network)이 도입되어 음성 향상 기법에서 우수한 성능을 내고 있다. 심화신경망을 이용한 음성 향상 기법은 다수의 은닉 층과 은닉 노드들을 통하여 잡음이 존재하는 음성 신호와 잡음이 존재하지 않는 깨끗한 음성 신호 사이의 비선형적인 관계를 잘 모델링하였다. 이러한 심화신경망 기반의 음성향상기법을 향상 시킬 수 있는 방법 중 하나인 강화학습을 적용하여 기존 심화신경망 대비 성능을 향상시켰다. 강화학습이란 대표적으로 구글의 알파고에 적용된 기술로써 특정 state에서 최고의 reward를 받기 위해 어떠한 policy를 통한 action을 취해서 다음 state로 나아갈지를 매우 많은 경우에 대해 학습을 통해 최적의 action을 선택할 수 있도록 학습하는 방법을 말한다. 본 논문에서는 composite measure를 기반으로 reward를 설계하여 기존 PESQ (Perceptual Evaluation of Speech



(그림 1) 제안하는 알고리즘의 블록도

Quality) 기반의 reward를 설계한 기술을 제안한다.

2. 강화학습 기반의 음성향상기법

강화학습이란 대표적으로 구글의 알파고에 적용된 기술로써 특정 state에서 최고의 reward를 받기 위해 어떠한 policy를 통한 action을 취해서 다음 state로 나아갈지를 매우 많은 경우에 대해 학습을 통해 최적의 action을 선택할 수 있도록 학습하는 방법을 말한다 [1], [2].

음성인식율과 상관관계가 더 높은 composite measure를 기반으로 reward를 설계하여 강화학습을 적용한다. 기존의 문헌에서 composite measure와 음성인식율과의 상관관계가 높다는 것은 많은 곳에서 입증되어 왔다. 이 입증된 composite measure를 적용하여 기존 PESQ기반 reward로 설계한 기술대비 높은 인식률을 얻을 수 있다.

강화학습이란 대표적으로 구글의 알파고에 적용된 기술로써 특정 state에서 최고의 reward를 받기 위해 어떠한 policy를 통한 action을 취해서 다음 state로 나아갈지를 매우 많은 경우에 대해 학습을 통해 최적의 action을 선택할 수 있도록 학습하는 방법을 말한다.

reward를 통해 학습한 action-value 함수를 통해 최적의 이득값 selection policy를 얻은 후 해당 이득값을 이용해 향상시킨 음성신호의 composite measure를 통해 reward를 계산한다. 이때 얻은 reward를 이용한 강화학습을 통해 최적의 selection policy 업데이트를 반복한다.

여기서 composite measure, Covl는 아래와 같다.

$$\text{Covl} = 1.594 + 0.805\text{SPESQ} - 0.512\text{SLLR} - 0.007\text{SWSS} \quad (1)$$

SPESQ, SLLR, SWSS는 각각 PESQ, log likelihood ratio, weighted spectral slope를 의미하는 수치이다.

단순히 composite measure는 성능뿐만 아니라 SNR, 잡음과 같은 외부적 요인에도 영향을 받기 때문에 절대적인 수치의 reward로 사용하는 것은 무리가 있다. 따라서 강화학습을 통해 음성향상을 시킨 결과와 대조군의 음성향상 결과사이의 composite measure 차이를 reward로 사용한다. 식으로 나타내면 아래와 같다.

$$R = \tanh(\alpha(Z - Z^{DNN})) \quad (2)$$

여기서 Z는 강화학습을 통해서 얻은 composite measure 값, Z^{DNN}는 대조군의 composite measure 값을 나타낸다.

또한 composite measure는 long-term에서 계산되는 반면 시간-주파수 이득값은 short-term에서 결정되어야하기 때문에 아래와 같은 추가적인 처리가 필요하다.

우선 식 (3)과 같이 계산을 통해 구하고자 하는 신호와의 차이를 구한다, 여기서 $Y_{w,k}$는 음성향상시킨 음성신호, $H_w S_{w,k}$는 깨끗한 음성신호를 의미한다.

$$\tilde{E}_k = \sum_{w=1}^{\Omega} |\ln|Y_{w,k}| - \ln|H_w S_{w,k}||^2 \quad (3)$$

식 (3)을 계산한 후 정규화를 시키는 식 (4)를 수행한다.

$$E_k = \frac{\tilde{E}_k}{\max_{k \in K}(\tilde{E}_k)} \quad (4)$$

식 (4)와 R을 기반으로 식 (5)와 같은 short-term의 reward를 계산한다.

$$r_k = \begin{cases} (1 - E_k)R & (R > 0) \\ E_k R & (\text{그외}) \end{cases} \quad (5)$$

새로운 reward, r_k를 이용하여 최적의 selection policy $\tilde{Q}$를 식 (6)과 같이 계산한다.

$$\widetilde{Q}(x_k,a_k)=\begin{cases} r_k+\max_{a\in A}Q(x_k,a) & (R>0) \\ Q(x_k,a_k) & (\text{그외}) \end{cases} \quad (6)$$

식 (6)을 살펴보면 강화학습을 통해 얻은 시간-주파수 이득값을 통해 얻은 음성향상 결과가 더 좋을 경우 selection policy에 r_k만큼의 reward를 주고 그렇지 않은 경우에는 reward를 주지 않은 것으로 생각할 수 있다.

훈련 데이터셋으로부터 대량의 $\widetilde{Q}$를 얻은 후, 강화학습의 출력인 Q와 최소가 되는 방향으로 action-value 함수를 학습한다. 수식으로 나타내면 아래와 같다.

$$\Theta_q \leftarrow \arg_\Theta \min \frac{1}{K}\sum_{k=1}^{K}\sum_{i=1}^{A}|\widetilde{Q}(x_k,i)-Q(x_k,i)|^2 \quad (7)$$

3. 실험 및 결과

학습에 사용된 데이터셋은 SiTEC이라는 DB에서 2000개의 문장을 사용했으며 고려된 잡음은 16개로 다음과 같다. babble, white, factory1, hfchannel, music, vacuum, water, animal, cough, door, footstep, refrigerator, shower, snore, TV, wind. 고려된 신호대잡음비는 0, 5, 10, 15, 20, 25 dB 이며 총 2000 x 16 x 6 = 192000개의 문장을 사용한 것이 된다.

테스트 셋은 SiTEC DB에서 학습에 사용되지 않은 1050개의 문장을 사용했으며 고려된 잡음은 다음과 같고 babble, factory, hfchannel, music, vacuum, water, refrigerator, TV, white, 고려된 신호대잡음비는 5, 10, 15, 20dB 이다.

비교모델은 적층형 IRM (ideal ratio mask) 기반의 심화신경망인데 구체적인 구조는 아래와 같다.

밑단의 입력은 잡음이 섞인 LPS이고 고려되는 프레임은 이전 5개, 현재, 미래 5개 총 11프레임이다. 윗단의 입력은 밑단 심화신경망을 통해 음성 향상된 신호의 LPS와 잡음이 섞인 신호의 LPS를 연결시켰으며 이전 4개, 현재, 미래 4개의 프레임을 고려해서 총 9개가 고려되었다. 타겟은 깨끗한 신호의 LPS이며 257차원이다.

강화학습의 구조는 257 차원의 잡음이 섞인 LPS 은닉층의 개수는 3개 은닉 노드의 개수는 각각 1024이며 출력은 32 차원이다. 활성함수는 ReLU, cost function은 MSE를 사용했으며 reward는 composite measure를 사용하였다.

고려되는 각 잡음환경별 인식률은 다음과 같다. 각 수치는 word error rate로써 낮을수록 더 좋은 인식율 성능을 나타내며, 구성한 인식기는 칼디기반이다.

4. 결론

기존 종래기술에서 사용된 측정방법보다 음성인식율과 보다 더 높은 상관관계를 갖는 composite measure를 기반으로 reward를 설계하고 이를 바탕으로 강화학습 기반의 음성향상기법을 제안하였다. 음성인식의 선행 과정으로 수행되어 음성 인식의 성능을 높일 뿐만 아니라 다양한 음성 통신 기술에 적용되어 음질을 향상시킬 수 있을 것이다.

<표 1> 여러 가지 잡음 환경에서의 WER (word error rate)

Noise Type	Method	SNR(dB)		
		5	10	15
Babble	대조군	64	40	24
	Proposed	61	38	23
Factory1	대조군	63	34	25
	Proposed	60	62	22
Hfchannel	대조군	50	30	23
	Proposed	49	28	23
Music	대조군	32	23	20
	Proposed	30	22	20
Vacuum	대조군	40	25	20
	Proposed	37	23	20
Water	대조군	55	34	25
	Proposed	54	33	25
Refrigerator	대조군	21	19	18
	Proposed	21	19	18
TV	대조군	44	26	21
	Proposed	42	24	21
White	대조군	73	34	22
	Proposed	71	33	22

감사의 글

본 연구는 2018년도 산업통상자원부 및 산업기술평가관리원(KEIT) 연구비 지원에 의한 연구임(10076583).

참고문헌

[1] Y. Koizumi, K. Niwa, Y. Hioka, K. Kobayashi, and Y. Haneda, “DNN-based source enhancement self-optimized by reinforcement learning using sound quality measurements” in *Proc.ICASSP*, March 2017, pp. 81 - 85.

[2] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou,H. King, D. Kumaran, D. Wierstra, S. Legg and D. Hassabis, “Human-level control through deep reinforcement learning,” *Nature*, 518, pp. 529 - 533, 2015.