음성 신호 전송을 위한 개선된 의미론적 통신 시스템

여예린, 김정현, 송홍엽

순천향대학교, 세종대학교, 연세대학교

yealin0817@gmail.com, j.kim@sejong.ac.kr, hysong@yonsei.ac.kr

An improved semantic communication system for speech transmission

Yerin Yeo, Junghyun Kim, Hong-Yeop Song

Soonchunhyang Univ., Sejong Univ., Yonsei Univ.

요 약

본 논문은 음성 데이터의 의미 정보를 전송하여 전송 효율을 향상 시키는 의미론적 통신시스템 모델에 대해 다룬다. 기존에 제안된 DeepSC-S-SER 모델은 좋은 성능을 보이지만 Channel Attenion만을 고려하고, 많은 학습파라미터로 인해 높은 복잡도를 갖는다. 따라서 CBAM-ResNet과 ECA-ResNet 모듈을 활용하여 성능 열화 없이 기존 모델의 한계를 극복하는 새로운 모델을 제안한다.

Ⅰ. 서 론

딥러닝이 다양한 분야에서 높은 성능을 달성하면서 통신 분야에서도 딥러닝을 적용한 연구가 많아졌다. 기존의 통신 시스템은 비트 또는 심볼 수준에서의 성능 개선에 중점을 두지만, Shannon과 Weaver를 통해 의미 수준에서의 정보를 전송하여 시스템의 효율성을 높일 수 있음이 나타났다. 정보와 사실을 모두 포함하는 개념인 의미론을 통해 데이터의 의미와 진실성을 고려할 수 있고, 이를 기반으로 의미 정보 간의 차이를 활용하는 의미론적 통신 시스템[1, 4]이 주목받고 있다.

의미론적 통신 시스템은 데이터의 의미 정보를 추출하여 전송하는 것을 주된 목표로 한다. 최근 이를 위해 Transformer 구조[2]를 기반으로 한 모델에 Squeeze and Excitation (SE) network[3]를 적용하여 음성 신호의 필수적인 정보와 특징을 학습 및 추출하는 모델인 DeepSC-S-SER[4]이 제안되었다. 이 모델은 좋은 성능을 보이지만 많은 학습 파라미터로 인해 높은 복잡도를 갖는다. 이를 해결하기 위해 최근 저복잡도 모델인 DeepSC-S-ECAR[5]이 제안되었다. 본 논문에서는 DeepSC-S-SER 모델의 성능에 근접하면서 Channel Attention뿐만 아니라 Spatial Attention까지 고려하기 위해 Convolutional Block Attention Module (CBAM)[6]을 적용한 DeepSC-S-CBAMR 모델과 DeepSC-S-ECAR의 복잡도를 더욱 낮춘 개선된 모델 DeepSC-S-IECAR를 제안한다.

Ⅱ. 본론

본 연구에서 사용한 데이터는 16KHz로 샘플링된 Edinburgh DataShare의 음성 데이터 세트이다. 전통적인 전화 시스템에서의 일반적인 음성 신호 샘플링 속도에 맞춰 데이터를 8KHz로 다운 샘플링한 후, 각 wav 파일의 음성 샘플 길이를 일치하게 만들고 훈련을 위해 프레임 변환을 하였다. 일반적으로 음성 데이터를 다룰 때, 진폭과 주파수 같은 음성 신호의 고유한 특징을 파악하기 위한 푸리에 변환, 스펙트럼, 스펙토그램 등의 전처리 작업을 거친다. 하지만 원시 음성 신호 자체에서 얻을 수 있는 또 다른 의미 정보가 존재할 수 있고 전처리 과정에서 생기는 시간 비용 등의 손실이 생길 수 있다. 이러한 이유로 본 논문에서는 진폭과 주파수 이외의 중요한 특징까지 고려할 수 있는 원시 데이터 사용하여 실험을 진행하였다.

그림 1은 음성 신호 전송을 위한 의미론적 통신 시스템의 구조이다. 시스템의 Semantic Encoder와 Semantic Decoder에서 데이터의 필수적인 정보를 학습하고 추출하도록 하는 Attention 기반 모듈을 사용하였고, 각 모델에 사용한 3개의 모듈 모두 기울기 소실 문제를 완화하기 위해 Residual Network (ResNet)에 적용하였다. 먼저 기존에 제안되었던 모델 DeepSC-S-SER에서는 Attention 기반 모듈 SE-ResNet 6개를 배치하였고, SE-ResNet에서 사용한 attention layer인 SE Layer는 2개의 Fully Connected (FC) 층으로 구성되어 있다.

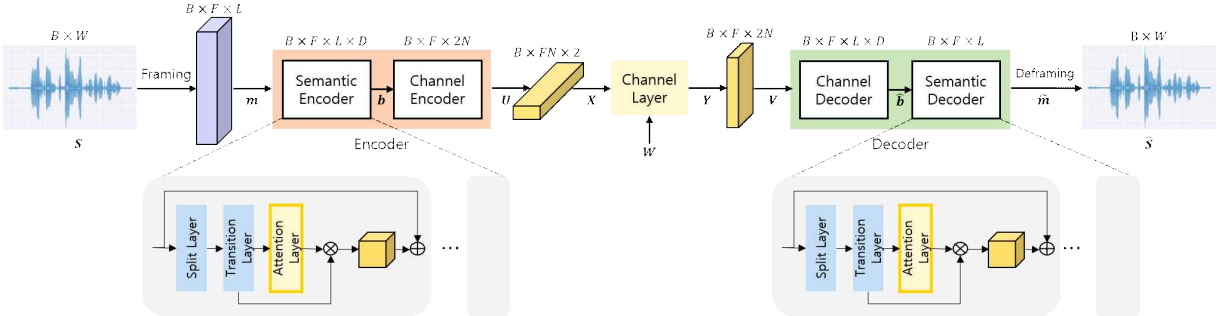

그림 1 음성 신호 전송을 위한 의미론적 통신 시스템

DeepSC-S-SER의 복잡도를 낮추기 위해 제안된 DeepSC-S-ECAR에서는 ECA-ResNet 모듈을 4개 배치하였다. ECA-ResNet의 ECA Layer[7]는 1D Convolution을 사용함으로써 지역적 특성을 효과적으로 학습할 수 있다. 또한 SE Layer와 다르게 FC 층을 사용하지 않아, 차원 축소 없이 Channel Attention을 수행하여 음성 신호의 전체적 특징을 효율적으로 포착하고 복잡도도 낮출 수 있다. 본 논문에서는 이를 모듈 수 3개로 개선한 DeepSC-S-IECAR을 제안하였다.

그림 2에 표현되어있는 CBAM-ResNet은 DeepSC-S-CBAMR에서 사용한 모듈이다. 6개의 모듈을 배치해서 사용하였고, CBAM-ResNet의 CBAM Layer는 Channel Attention 모듈과 Spatial Attention 모듈로 구성된다. 먼저, Channel Attention 모듈에서는 채널 간의 관계를 살펴 어떤 채널에 더 집중할지 인코딩하고, Spatial Attention 모듈에서는 전체 채널 내 픽셀 중 어디에 더욱 집중할 것인지 인코딩한다. Spatial Attention까지 고려함으로써 Channel Attention만 고려하는 SE Layer보다 특징에 대한 집중도를 높일 수 있다.

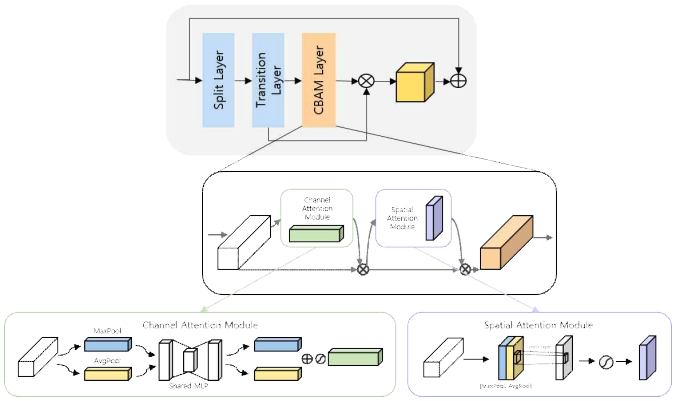

그림 2 CBAM-ResNet의 구조

의미론적 통신시스템은 의미를 복원하는 것이 목표이므로 원본 음성 신호와 복원된 음성 신호 사이의 왜곡을 측정하는 signal to distortion ration (SDR)[8]을 성능 지표로 채택하였다. 성능 비교를 위해 배치 크기는 8개, 학습 반복 횟수는 100회로 설정하고 Adam 최적화기를 학습률 0.001로 훈련을 진행하였다. 그림 5는 기존 모델인 DeepSC-S-SER과 제안하는 모델인 DeepSC-S-CBAMR 및 DeepSC-S-IECAR의 성능을 비교한 것이다. 각 SNR에서 성능 열화가 거의 없는 것을 확인할 수 있다.

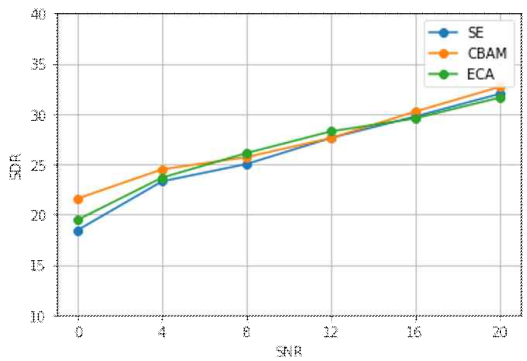

그림 3 SNR에 따른 모델별 성능 비교

표 1은 두 기존 모델과 제안하는 두 모델의 파라미터 수를 비교한 것이다. 기존 모델 대비 제안하는 모델 DeepSC-S-IECAR의 파라미터 수가 약 48% 감소한 것을 확인하였다. 특히 ECA를 사용한 저복잡도 모델 DeepSC-S-ECAR과 비교해도 파라미터 수가 약 24% 감소하였다.

표 1. 모델별 파라미터 수 비교

	파라미터 수
DeepSC-S-SER [4]	344,179
DeepSC-S-ECAR [5]	233,843
DeepSC-S-CBAMR	342,715
DeepSC-S-IECAR	179,443

Ⅲ. 결론

본 논문에서는 음성 데이터의 의미 정보를 전송하여 전송 효율을 향상시키는 의미론적 통신시스템 모델 DeepSC-S-SER의 한계를 극복하기 위한 DeepSC-S-CBAMR 모델과, 저복잡도 모델 DeepSC-S-ECAR의 복잡도를 더욱 낮춘 개선된 모델 DeepSC-S-IECAR을 제안한다. 제안한 모델 모두 기존 모델 대비 성능 열화가 거의 없었으며, 특히 DeepSC-S-IECAR 모델은 복잡도에 영향을 미치는 파라미터 수를 약 48% 감소시켰다. 이러한 결과를 바탕으로 향후 음성 신호 뿐만 아니라 이미지, 비디오 등 다양한 분야에 적용할 수 있을 것으로 기대된다.

ACKNOWLEDGMENT

이 (성과)는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No.2020R1A2C2011969).

참 고 문 헌

[1] H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, “Deep learning enabled semantic communication systems,” IEEE Trans. Signal Process., pp. 1 - 1, Apr. 2021.

[2] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez,u. Kaiser, and I. Polosukhin, “Attention is all you need,” in Proc. 31stInt. Conf. Neural Inf. Process. Syst. (NIPS), Long Beach, CA, USA,Dec. 2017, pp. 6000 - 6010.

[3] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, “Squeeze-and-excitation networks,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 42, no. 8, pp. 2011 - 2023, Aug. 2020.

[4] Z. Weng and Z. Qin, “Semantic communication systems for speech transmission,” IEEE J. Sel. Areas Commun., vol. 39,no. 8, pp. 2434 - 2444, Aug. 2021.

[5] 여예린, 김정현, 송홍엽. “음성 신호 전송을 위한 저복잡도의 의미론적 통신시스템,” 한국통신학회 추계종합학술발표회, Nov, 2022.

[6] WOO, Sanghyun, et al. Cbam: Convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV). p. 3-19, 2018.

[7] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks", Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. , pp. 11534-11542, 2020.

[8] E. Vincent, R. Gribonval, and C. Fevotte, “Performance measurement ´in blind audio source separation,” IEEE Trans. Audio, Speech, Lang.Process., vol. 14, no. 4, pp. 1462 - 1469, Jul. 2006.