음성 신호 전송을 위한 저복잡도의 의미론적 통신 시스템

여예린, 김정현, 송홍엽

순천향대학교, 세종대학교, 연세대학교

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

A low-complexity semantic communication system for speech transmission

Yerin Yeo, Junghyun Kim, Hong-Yeop Song

Soonchunghyang Univ., Sejong Univ., Yonsei Univ.

요 약

본 논문은 통신 시스템에서 음성 신호의 의미적인 정보를 효율적으로 복구하는 모델에 대해서 다룬다. SE-ResNet 모듈을 사용하는 기존 DeepSC-S-SER 모델은 좋은 성능을 갖지만 많은 학습 파라미터가 존재하여 높은 복잡도를 갖는다. 따라서 ECA-ResNet 모듈을 활용하여 적은 수의 학습 파라미터로 기존 모델의 성능에 근접하는 새로운 모델을 제안한다.

Ⅰ. 서 론

딥러닝이 다양한 분야에서 높은 성능을 달성하면서 통신 분야에서도 딥러닝을 적용한 연구[1-2]가 많아졌다. 기존의 통신 시스템은 비트 또는 심볼 수준에서의 성능 개선에 중점을 두지만, Shannon과 Weaver를 통해[3] 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[4-7]이 주목받고 있다.

의미론적 통신 시스템은 데이터의 의미 정보를 추출하여 전송하는 것을 주된 목표로 한다. 최근 의미론적 통신 시스템을 위해 Transformer 구조[8]를 기반으로 한 모델에 Squeeze and Excitation (SE) network[9]를 적용하여 음성 신호의 필수적인 정보와 특징을 학습 및 추출한 후 신호 복구에 활용하는 DeepSC-S-SER 모델[10]이 제안되었다. 본 논문에서는 DeepSC-S-SER 모델의 성능에 근접하면서 Efficient Channel Attention (ECA) network [11]를 적용하여 복잡도를 현저히 낮춘 DeepSC-S-ECAR을 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 샘플링 한 후, 각 wav 파일의 음성 샘플 수를 일치하게 만들기 위해 입력 샘플의 길이를 고정하고 훈련을 위해 프레임 변환을 하였다.

그림 1은 음성 신호 전송을 위한 의미론적 통신 시스템의 구조이다. 기존 모델인 DeepSC-S-SER와 제안하는 모델 DeepSC-S-SER은 시스템 구조의 Semantic Encoder와 Semantic Decoder에서 데이터의 필수적인 정보를 학습하고 추출하도록 하는 Attention 기반 모듈인 SE-ResNet과 ECA-ResNet을 각각 사용하였다.

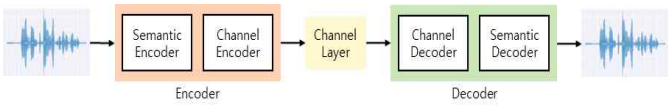

그림 1. 음성 신호 전송을 위한 의미론적 통신 시스템

이때 사용된 SE-ResNet과 ECA-ResNet은 SENet과 ECANet을 Residual Network (ResNet)에 적용한 모듈이다. 기존 모델은 그림 2에 표현된 SE-ResNet 6개를 사용하고, SE-ResNet의 SE Block은 2개의 Fully Connected (FC) 층으로 구성되어 있다. 반면, 제안하는 모델은 그림 3에 표현된 ECA-ResNet 4개를 적용한다. ECA-Net의 ECA Block은 2개의 FC 층 대신 1D Convolution을 사용함으로써 지역적 특성을 효과적으로 학습할 수 있고 차원 축소 없이 channel attention을 수행하여 음성 신호의 전체적 특징을 효율적으로 포착하여 복잡도를 낮출 수 있다.

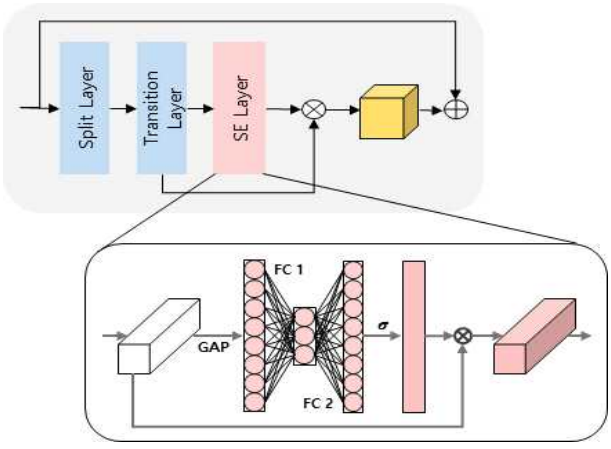

그림 2. SE-ResNet의 구조

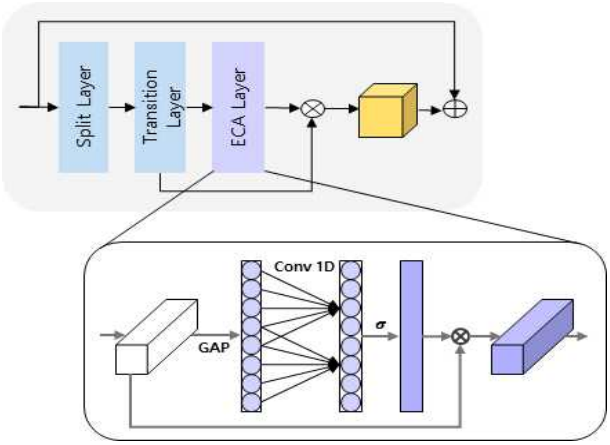

그림 3. ECA-ResNet의 구조

딥러닝 기반 통신 시스템은 비트로 표현된 전송 메시지를 정확하고 효율적으로 복구하는 것이 목표이므로 주로 bit error rate (BER) 또는 symbol error rate (SER)을 성능 지표로 사용한다. 반면 제안하는 모델은 의미를 복원하는 것에 초점을 맞추므로 본 논문에서는 원본 음성 신호와 복원된 음성 신호 사이의 전체적인 오류를 측정하는 signal to distortion ration (SDR)[12]을 성능 지표로 채택하였다. 성능 비교를 위해 배치 크기는 8개, 학습 반복 횟수는 10회로 설정하고 Adam 최적화기를 학습률 0.001로 설정하여 훈련을 진행하였다. 그림 4는 기존 모델인 DeepSC-S-SER와 제안하는 모델인 DeepSC-S-ECAR의 성능을 비교한 것이다. 각 SNR에서의 성능 차이가 거의 없는 것을 확인할 수 있었으며 일부 SNR에서는 더 높은 성능을 보였다.

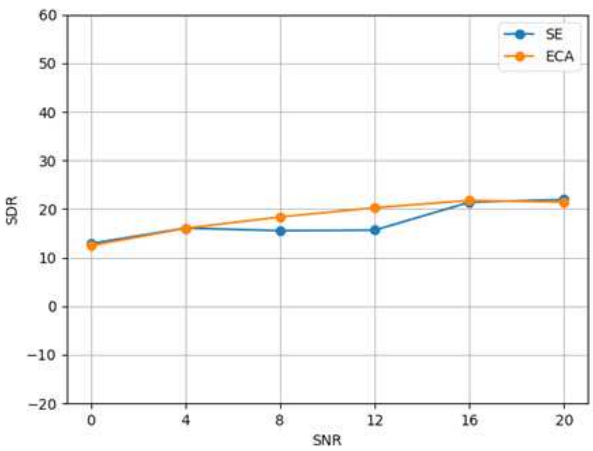

그림 4. 기존 DeepSC-S-SER[10]과 제안 모델의 성능 비교

표 1은 기존 모델과 제안하는 모델의 파라미터 수를 비교한 것이다. 기존 모델 대비 제안하는 모델의 파라미터 수가 약 32% 감소한 것을 확인하였다.

표 1. 기존 DeepSC-S-SER과 제안 모델의 파라미터 수 비교

	파라미터 수
DeepSC-S-SER [10]	344,179
DeepSC-S-ECAR	233,843

Ⅲ. 결론

본 논문에서는 음성 데이터의 의미 정보를 전송하여 전송 효율을 향상시키는 DeepSC-S-SER를 복잡도 측면에서 개선한 DeepSC-S-ECAR를 제안하였다. 제안한 모델은 기존 모델 대비 성능 열화가 거의 없었으며, 복잡도에 영향을 미치는 파라미터 수를 약 32% 감소시켰다. 이러한 결과를 바탕으로 향후 복잡도뿐만 아니라 성능을 개선하는 새로운 모델 설계도 가능할 것으로 기대된다.

ACKNOWLEDGMENT

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

참 고 문 헌

[1] Z. Qin, H. Ye, G. Y. Li, and B.-H. F. Juang, “Deep learning in physicallayer communications,” IEEE Wireless Commun., vol. 26, no. 2, pp.93-99, Apr. 2019.

[2] Z. Qin, G. Y. Li, and H. Ye, “Federated learning and wireless communications,” arXiv preprint arXiv:2005.05265, pp.134-140, May. 2020.

[3] C. E. Shannon and W. Weaver, The Mathematical Theory of Communications., The University of Illinois Press, 1949.

[4] 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.

[5] Z. Weng, Z. Qin, and G. Y. Li, “Semantic communications for speech signals,” in Proc. IEEE Int. Conf. Commun. (ICC), Montreal, QC,Canada, pp.1-6, Jun. 2021.

[6] H. Xie and Z. Qin, “A lite distributed semantic communication system for Internet of Things,” IEEE J. Sel. Areas Commun., vol. 39, no. 1, pp. 142-153, Jan. 2021.

[7] B. Güler, A. Yener, and A. Swami, “The semantic communication game,” IEEE Trans. Cogn. Commun. Netw., vol. 4, no. 4, pp. 787-802, Dec. 2018.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.