Autoencoder-based Data Compression Model Experiment for Semantic Communication

Jinyoung Oh
Department of Software
Kongju National University
Cheonan, Korea
wlsdud6221@naver.com

Yunkyung Choi
Department of Software
Kongju National University
Cheonan, Korea
chldbsrud010@naver.com

Chanyoung Oh
Department of Software
Kongju National University
Cheonan, Korea
cyoh@kongju.ac.kr

Woongsoo Na
Department of Software
Kongju National University
Cheonan, Korea
wsna@kongju.ac.kr

***Abstract*—In recent years, the convergence of neural network architectures and semantic communication has led to innovative strides in representation learning. This paper explores the application of autoencoders, a subset of neural networks designed for unsupervised learning, in encoding and decoding data to capture semantic nuances. Additionally, we discuss future research directions, proposing a semantic communication model learned from time series data and presenting experimental results. Our experiments involve encoding and decoding time series data using autoencoders, evaluating the feasibility of integrating autoencoder technology into semantic communication. Following the experiment, our proposed model exhibited a loss rate of approximately 0.15-0.2% for the time series data. This outcome represents a notable result, especially when compared to the compression rate of transmission data. It hints at the potential for a future autoencoder-based semantic communication model.**

***Index Terms*—Semantic communication, autoencoder, encoding/decoding, and data compression.**

I. Introduction

In recent years, the intersection of neural network architectures and semantic communication has paved the way for innovative approaches in representation learning. One notable paradigm within this domain is the application of autoencoders, a class of neural networks designed for unsupervised learning, to encode and decode data in a manner that captures the semantic essence of the input. The amalgamation of autoencoder architectures with the imperative of semantic communication holds promise for transforming raw data into compact, meaningful representations, fostering advancements in various fields such as image processing, natural language understanding, and anomaly detection [1].

Autoencoders, comprising an encoder and decoder, have demonstrated efficacy in transforming high-dimensional input data into a lower-dimensional latent space, thereby encapsulating essential features while discarding noise. This latent space, by virtue of semantic communication, encapsulates not only the structural attributes of the input but also the underlying meaning, contributing to the interpretability and applicability of the learned representations. The endeavor to imbue neural networks with semantic understanding is particularly critical in tasks where capturing meaningful patterns and discarding extraneous details is paramount.

This paper delves into the landscape of autoencoder-based semantic communication technology, exploring the foundational concepts, architectural variations, and applications that underscore its significance in representation learning [2]–[6]. We navigate through the intricate interplay between autoencoders and semantic communication, shedding light on how this synergy empowers the network to learn representations that transcend mere reconstruction, delving into the realm of semantic richness. The subsequent sections of this paper will elaborate on the components of autoencoder-based semantic communication, detailing the encoder-decoder framework, the latent space, and the nuanced challenges associated with preserving semantic information during the learning process. Additionally, we will survey the diverse applications across domains such as image and video processing, natural language understanding, and anomaly detection, where the marriage of autoencoders and semantic communication has exhibited transformative potential. Finally, we discuss future semantic communication research by proposing a semantic communication model learned from time series data and presenting experimental results.

II. Structural Model of Autoencoder-Based Semantic Communication

Autoencoders, as a fundamental architecture in neural network-based representation learning, offer a versatile framework for encoding and decoding input data. In the realm of semantic communication, the structural model of an autoencoder undergoes crucial adaptations to ensure that the learned representations not only encapsulate the structural intricacies of the input but also convey the semantic essence embedded within the data.

A. Encoder: Unveiling Semantic Features

The encoder, serving as the entry point of the autoencoder, is meticulously designed to unfold semantic features from the input data. Beyond its traditional role of extracting hierarchical

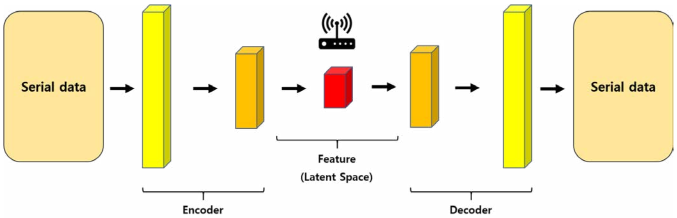

Fig. 1. Autoencoder-based semantic communication model structure.

representations, the encoder in the context of semantic communication is tailored to discern and encode the underlying meaning inherent in the input as shown in Fig. 1. This involves the exploration of neural network architectures that adeptly capture semantic content through learned features.

B. Latent Space: Semantics in Compression

The latent space, representing the compressed form of the input, becomes a focal point for semantic communication. It is imperative that the latent space not only encapsulates statistical regularities but also forms a cohesive semantic landscape. Proximity in the latent space should correspond to similarity in meaning, fostering a nuanced representation that transcends the mere replication of structural patterns.

C. Decoder: Recreating Meaningful Outputs

In the decoding phase, emphasis is placed on the recreation of meaningful outputs from the latent space. The decoder is tasked not only with reconstructing the input structure but also with generating outputs that convey the intended semantic information. This aligns with the overarching objective of semantic communication — ensuring that the reconstructed output encapsulates the underlying meaning of the input.

D. Training Objective: Balancing Structure and Semantics

The training objective is twofold: minimize reconstruction error to capture structural details and incorporate semantic constraints to guide the learning of meaningful features. This dual objective is crucial for achieving a balanced representation that not only faithfully reconstructs the input but also communicates its semantic content effectively. Regularization techniques and auxiliary tasks may be introduced to impart semantic constraints during training.

E. Semantic Constraints: Guiding the Learning Process

To enhance semantic communication, various constraints may be introduced, shaping the learning process towards capturing and preserving semantically relevant features. Techniques such as adversarial training, attention mechanisms, or the incorporation of semantic labels serve as levers to guide the network towards a more nuanced and meaningful representation in the latent space.

F. Hyperparameter Tuning: Fine-Tuning for Semantic Fidelity

The success of the structural model hinges on effective hyperparameter tuning. Configuring the number of layers, units per layer, learning rates, and other parameters is a nuanced process. Fine-tuning these hyperparameters ensures that the model strikes the right balance between capturing structural details and conveying semantic meaning.

III. Applications of Autoencoder-Based Semantic Communication Technology

Autoencoder-based semantic communication technology has witnessed widespread adoption across diverse domains, showcasing its adaptability and efficacy in transforming and comprehending data. This section provides an overview of key applications supported by literature in various fields:

- **Image Processing:** Autoencoders have proven instrumental in image processing tasks, including compression, denoising, and feature extraction. Research in computer vision and image processing conferences, highlights their role in enhancing tasks like facial recognition and object detection [2].
- **Natural Language Understanding:** In the realm of natural language processing, autoencoders contribute significantly to semantic understanding of textual data. In this domain, they are employed for tasks such as document summarization, sentiment analysis, and language translation by capturing the underlying meaning and structure of textual information [3].
- **Anomaly Detection:** Autoencoders excel in anomaly detection by learning the normal patterns within datasets. Any deviation from the learned normal behavior is flagged as an anomaly, making them effective in identifying outliers in various fields, including cybersecurity and fault detection in achinery. [4].
- **Time Series Analysis:** Autoencoders are utilized for encoding and decoding temporal patterns in time series data. They help in recognizing trends, predicting future values, and identifying anomalies in sequences of data, making them valuable in finance, healthcare, and environmental monitoring. [5].
- **Healthcare Imaging** Autoencoders play a role in medical image analysis by extracting meaningful features from medical images. They contribute to tasks like image segmentation, disease diagnosis, and anomaly detection in medical scans, improving the accuracy and efficiency of diagnostic processes. [6].

As previously noted, autoencoder-based applications have undergone diverse investigations, yet there is limited exploration of their application to semantic communication. Nonetheless, the capability to extract semantic data and transmit it via an autoencoder represents a technology poised to decrease overall network traffic.

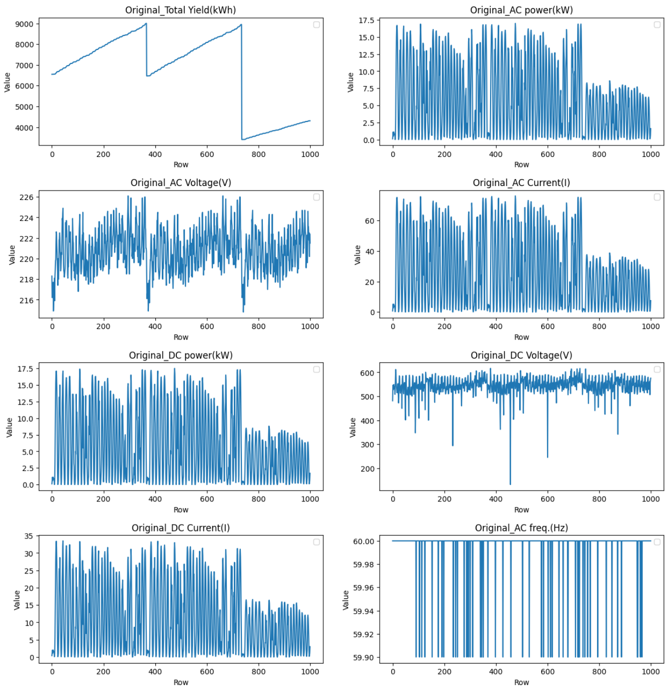

Fig. 2. Original solar data used in experiments.

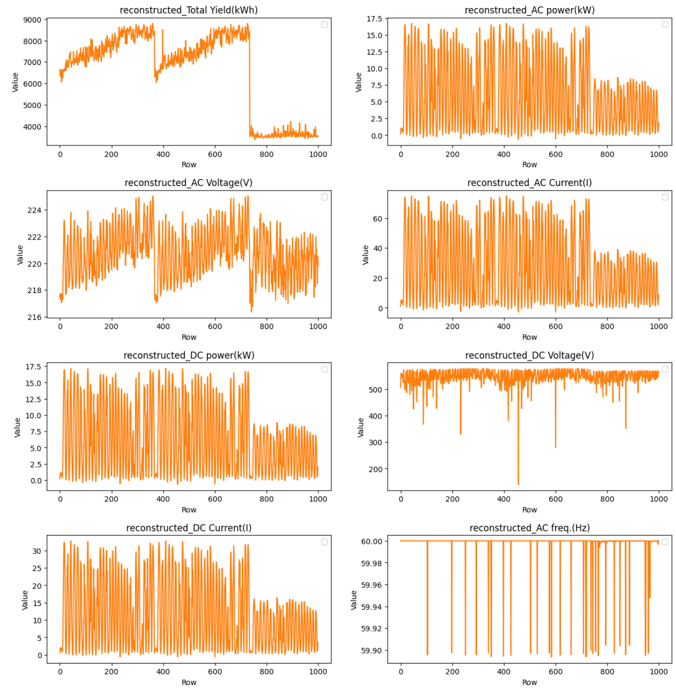

Fig. 3. Restored Solar data through autoencoder.

IV. Experimental Test: Autoencoder-based Semantic Communication

In this section, we present experimental findings involving the encoding/decoding of time series data using an autoencoder, aiming to assess the viability of incorporating autoencoder technology into semantic communication.

The dataset employed in this investigation pertains to solar heat and constitutes a time series dataset structured with information collected over distinct time periods. Comprising a total of 9 columns, the dataset includes 3 columns each for date, total power generation, frequency, and AC and DC information. The dataset encompasses over 100,000 rows, signifying the substantial volume of data available for analysis.

The implementation of the model took place within the Jupyter environment, utilizing the Keras library for generating the autoencoder model. The neural network was structured with both the encoder and decoder consisting of three dense layers each, employing the activation function *tanh*. To mitigate overfitting during the model's learning process, the optimizer was set to *Adam*, and the learning rate was adjusted arbitrarily to 0.001. The primary objective of this model is to compress 8-dimensional data into 2 dimensions, extract relevant feature values, and subsequently reconstruct the data back to its original 8-dimensional form.

Fig. 2 and Fig. 3 shows the original and restored solar data used in experiments, respectively. Given the varied data ranges in each column of this dataset, an initial step involves normalization through MinMaxScaler, mapping the values to a range between 0 and 1. This normalization is crucial to ensure accurate restoration by the model, considering the diverse value scales. Subsequently, the normalized data is fed into the neural network as input, undergoing compression to a lower dimension where essential feature elements are extracted. Following this compression, the data is restored to its original dimension, and the results are scrutinized for accuracy.

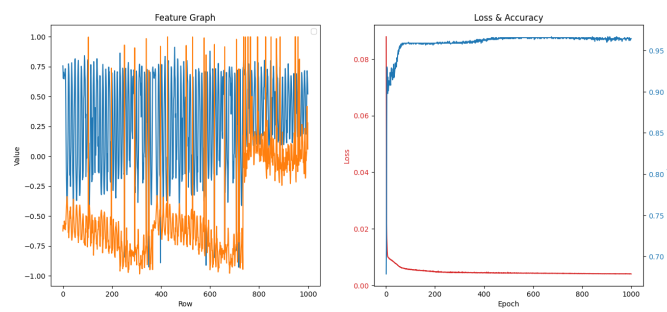

Fig. 4. Latent space (semantic information) and accuracy result.

Fig. 4 shows the latent space graph compressed into 2-dimentional data and accuracy. Especially, right side of Fig. 4 illustrates the loss and accuracy values for each epoch. In models not focused on classification, accuracy values may not carry significant meaning. Instead, the loss function serves as the foundational metric for assessing accuracy. In this particular model, Mean Squared Error (MSE) was employed as the loss function. After 1000 epochs and 1000 data points, the observed results ranged between 0.0015 and 0.0023. While the overall trend of the graph aligns with the original, a limitation arises wherein the restoration rate decreases, particularly in scenarios with minimal or absent changes in time series data.

V. CONCLUSION

This paper introduces the application of autoencoder structures as a semantic communication technique based on 6G standard technology. Additionally, it presents a model for data restoration through autoencoders, designed for feasibility experiments. The experiment showcased diverse outcomes contingent on the activation function and layers of the neural network. Notably, a high recovery rate was achieved with the *tanh* function. However, challenges surfaced, particularly in scenarios with minimal or non-existent changes, necessitating the creation of specialized models for both structured and unstructured time series data. The integration of communication methods like the presented model with upcoming 6G standard technology holds promise for rapid transmission rates and minimal data loss even in extensive data communications.

ACKNOWLEDGMENT

This work was supported by the Priority Research Centers Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2019R1A6A1A03032988) and in part by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2023- RS-2022-00156353) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation.

REFERENCES

[1] H. Xie, Z. Qin, G. Y. Li, and B. H. Juang. Deep learning enabled semantic communication systems. *IEEE Transactions on Signal Processing*, 69:2663–2675, 2021.

[2] S. Y. Wang, X. Enoch and D. D. Kim. Semantic preserving adversarial attack generation with autoencoder and genetic algorithm. *In GLOBECOM 2022-2022 IEEE Global Communications Conference*, pages 80–85, 2022.

[3] A. Joshi, E. Fidalgo, E. Alegre, and L. Fernández-Robles. Summcoder: An unsupervised framework for extractive text summarization based on deep auto-encoders. *Expert Systems with Applications*, 129:200–215, 2019.

[4] K. Jang, S. Hong, M. Kim, J. Na, and I. Moon. Adversarial autoencoder based feature learning for fault detection in industrial processes. *IEEE Transactions on Industrial Informatics*, 18(2):827–834, 2021.

[5] C. Yin, S. Zhang, J. Wang, and N. N. Xiong. Anomaly detection based on convolutional recurrent autoencoder for iot time series. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 52(1):112–122, 2020.

[6] J. C. Kim and K. Chung. Multi-modal stacked denoising autoencoder for handling missing data in healthcare big data. *IEEE Access*, 8:104933–104943, 2020.