Meeting Note

Author: Henry Lin
Date: 2024/07/22 – 2024/07/26

Summary
Time Series Data Decomposition. [1]
LLM Enhanced Classification Tasks.
Text Dense Retrieval. [2]
Speech Recognition Pre-train. [3]

Plan for Next Week
LLM Enhanced model.
Reinforcement Learning on LLM.

Details
Zeiler et al. [1] introduced a method to decompose time series data using Empirical Mode Decomposition (EMD). We first identify the mean envelope (M) by determining the local minima and maxima of the original data, then subtract this mean envelope from the original data to obtain the first Intrinsic Mode Function (IMF1). After obtaining IMF1, we treat the residual data as the new original data and repeat the process to obtain subsequent IMFs (IMF2, IMF3, etc.). As the iteration proceeds, the IMF will look smoother.

When machine learning models perform classification tasks, they can sometimes produce unexpected or inaccurate results. For instance, an OCR translation model might encounter the text "지방" in the context of nutrition facts and translate it as "region" instead of "fat." While "지방" can mean either "region" or "fat" depending on the context, "fat" is the more appropriate translation in this scenario. Similarly, in a multi-label image classification task, the model might return a label set such as "mountain, horse, grass, traffic light." Although this set of labels might be accurate in a very unusual image, the label "traffic light" is likely to be a misprediction in most cases.
To address this problem, Large Language Models (LLMs) may prove to be useful. LLMs have the ability to consider contextual information, enabling them to filter out unreasonable outputs. By leveraging their contextual understanding,

LLMs can enhance the accuracy and relevance of the classifications and translations provided by machine learning models, even the speech recognition can also be enhanced by LLMs, for example, by fixing homophones.

Guo et al. [2] introduced a novel architecture for the text dense retrieval. This model uses the Transformer's encoders to encode the text into a representation without using decoders. The traditional architectures typically pair an encoder and a decoder; however, the decoder may have the bypass effect on the encoder, which will make the decoder to bypass the dependency on the output of the encoder. The key idea is to force the encoder to generate text representations close to its own random spans while keeping them far away from others using a group-wise contrastive loss.

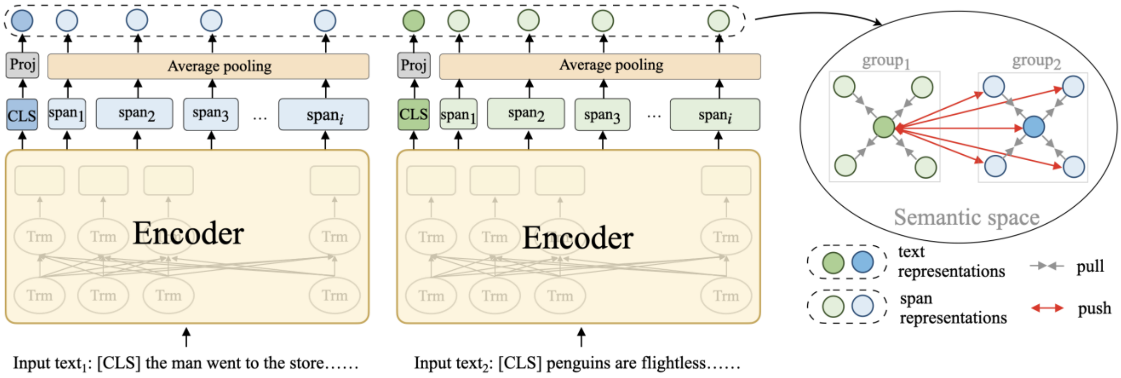

Figure 1. COSTA architecture.

Chiu et al. [3] introduced the BERT-based Speed pre-Training with Random-projection Quantizer (BEST-RQ). They first mask the audio data randomly, then feed it into an ASR encoder, and the unmasked audio is feed into a frozen projector and a randomly-initialized module called "codebook" which performs a nearest-neighbor lookup. And the objective of the ASR encoder is to predict which labels are corresponding to the masked signals.

This approach provides flexibility and compatibility since the random-projection quantizer doesn't need any training and is separated from the main architecture. This study shows the great performance of random-projection quantization for speech recognition without relying on labeled data.

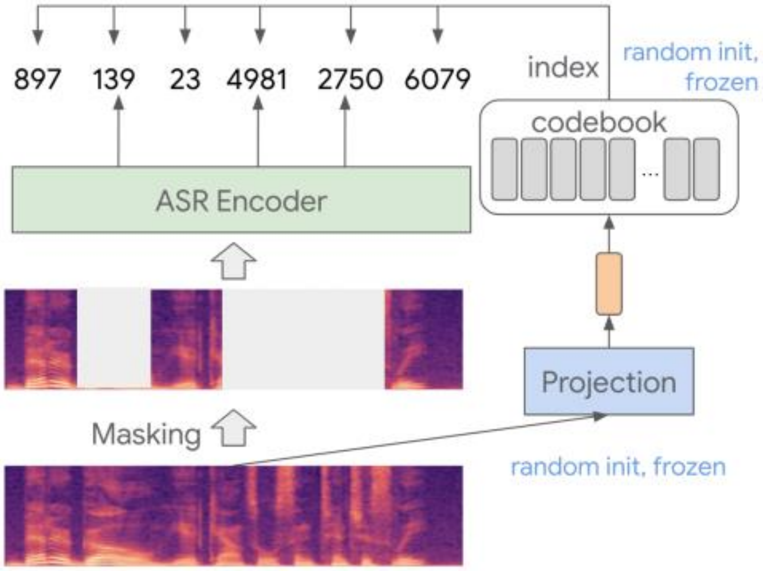

Figure 2. BEST-RQ architecture.

References

[1] Zeiler, A., Faltermeier, R., Keck, I. R., Tomé, A. M., Puntonet, C. G., & Lang, E. W. (2010, July). Empirical mode decomposition-an introduction. In The 2010 International Joint Conference on Neural Networks (IJCNN) (pp. 1–8). IEEE.

[2] Xinyu Ma, Jiafeng Guo, Ruqing Zhang, Yixing Fan and Xueqi Cheng. 2022. Pre-train a Discriminative Text Encoder for Dense Retrieval via Contrastive Span Prediction. In Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '22), July 11–15, 2022, Madrid, Spain. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3477495.3531772

[3] CHIU, Chung-Cheng, et al. Self-supervised learning with random-projection quantizer for speech recognition. In: International Conference on Machine Learning. PMLR, 2022. p. 3915-3924.