

Peer Review

Review of: "k2SSL: A Faster and Better Framework for Self-Supervised Speech Representation Learning"

Murali Karthick Baskar¹

1. Independent researcher

The paper proposes to improve the HuBERT model with Zipformer architecture and change the optimizer in the K2 framework. The model is analysed using Librispeech, and the code is available as open-source.

The authors provide wide ablations to understand the importance of their changes. The results show that their proposed modifications lead to memory gains by fitting more speech samples in a single batch compared to the transformer, as in table 1. The performance improvements with unlabeled data and the n-gram LM in tables 4 and 3 show that the proposed architecture is complementary.

The results can be further justified by analysing the performance on other OOD datasets such as TEDlium.

Declarations

Potential competing interests: No potential competing interests to declare.