Qeios

Peer Review

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

Dr Deepak Mishra[1]

1. COMMUNICATION AND NAVIGATION PAYLOAD AREA, Indian Space Research Organization, Bengaluru, India

The paper presents two contributions:

1. k2SSL Framework: A toolkit offering computational advantages in self-supervised speech learning.
2. Zipformer-Based System: A model leveraging the Zipformer encoder with the ScaledAdam optimizer for enhanced learning performance.

While the computational benefits and performance improvements are compelling, the paper lacks clarity in distinguishing between the framework and the model. For instance, the last paragraph of Section I initially introduces k2SSL as a general SSL framework for ASR (Automatic Speech Recognition), but later conflates it with the Zipformer-specific implementation. Similarly, Table II suggests k2SSL as a general framework (or toolkit) which will allow custom models to be trained with the computational advantages as described in Section III.D, while the conclusion again melds k2SSL with the Zipformer backbone-based model.

Strengths:

1. Innovative contributions with practical implications.
2. Detailed comparisons showcasing computational efficiency.

Weaknesses:

1. Ambiguity in defining k2SSL and its generalizability.
2. Inconsistent messaging across sections.

doi.org/10.32388/M40YBZ

Recommendations:

1. Clearly differentiate k2SSL as a framework and the Zipformer backbone-based model as a specific use case.

In summary, while promising, the paper would benefit from improved clarity.

Declarations

Potential competing interests: No potential competing interests to declare.