Actes de CORIA-TALN 2026
Nantes, France, 29 juin au 3 juillet 2026

Modèle SENSE : une solution open source pour les tâches multilingues et multimodales basées sur la sémantique

Salima Mdhaffar[1] Haroun Elleuch[1, 2] Chaimae Chellaf[1,3] Maryem Bouziane[1]
Ha Nguyen[4] Yannick Estève[1]
(1) Avignon Université, LIA, France
(2) Elyadata, France
(3) Lundi Matin, France
(4) Oracle, France
salima.mdhaffar@univ-avignon.fr

RÉSUMÉ

Cet article présente SENSE (Shared Embedding for N-lingual Speech and tExt), une solution open source inspirée du cadre SAMU-XLSR et conceptuellement proche des modèles SONAR de Meta AI. Ces approches reposent sur un cadre teacher–student visant à aligner un encodeur de parole auto-supervisé avec les représentations continues indépendantes de la langue produites par un encodeur de texte, au niveau de l'énoncé. Nous décrivons comment la méthode originale SAMU-XLSR a été améliorée en sélectionnant un modèle texte enseignant plus performant ainsi qu'un meilleur encodeur de parole initial. Le code source permettant d'entraîner et d'utiliser les modèles SENSE a été intégré dans l'outil SpeechBrain [1], et le premier modèle SENSE que nous avons entraîné a été rendu public [2]. Nous présentons des résultats expérimentaux sur des tâches sémantiques multilingues et multimodales, dans lesquelles notre modèle SENSE atteint des performances très compétitives. Enfin, cette étude apporte de nouveaux éclairages sur la manière dont la sémantique est capturée dans ce type d'encodeurs de parole alignés sémantiquement.

ABSTRACT

SENSE models : an open source solution for multilingual and multimodal semantic-based tasks

This paper introduces SENSE (Shared Embedding for N-lingual Speech and tExt), an open-source solution inspired by the SAMU-XLSR framework and conceptually similar to Meta AI's SONAR models. These approaches rely on a teacher–student framework to align a self-supervised speech encoder with the language-agnostic continuous representations of a text encoder at the utterance level. We describe how the original SAMU-XLSR method has been updated by selecting a stronger teacher text model and a better initial speech encoder. The source code for training and using SENSE models has been integrated into the SpeechBrain toolkit, and the first SENSE model we trained has been publicly released. We report experimental results on multilingual and multimodal semantic tasks, where our SENSE model achieves highly competitive performance. Finally, this study offers new insights into how semantics are captured in such semantically aligned speech encoders.

MOTS-CLÉS : Encodeur de parole multilingue, représentation sémantique, recherche d'information multimodale, traduction de la parole.

KEYWORDS: multilingual speech encoder, semantic representation, multimodal information retrie-

1. https ://github.com/speechbrain/speechbrain/tree/develop/recipes/CommonVoice/SENSE
2. https ://huggingface.co/LIA-AvignonUniversity/SENSE/tree/main

Actes des 33ème Conférence sur le Traitement Automatique des Langues Naturelles.
Volume 2 : articles déjà publiés, pages 33–34.
Cette œuvre est mise à disposition sous licence Attribution 4.0 International.

val, speech translation.

ARTICLE ACCEPTÉ À : IEEE ASRU 2025 : Workshop on Automatic Speech Recognition and Understanding, Honolulu, Hi, USA, December, 6-10 2025..
URL : https://ieeexplore.ieee.org/abstract/document/11433845