

Context-aware multilingual semantic representations of dialog turns for SLU task



Contact

sahar.ghannay@lisn.upsaclay.fr, sophie.rosset@lisn.upsaclay.fr

Description

In the context of the spoken language understanding (SLU) field for dialogue systems, the problem of contextual representation remains a hot topic despite the many works on it [Tomashenko et al., 2020]. Focusing on this problem, the main objective of this study is to build a context-aware representation of dialog turns, enriched with multilingual multimodal semantic information.

A recent study [Laperrière et al., 2023] investigates a specific in-domain semantic enrichment of the SSL (self-supervised learning) SAMU-XLSR model by specializing it on a small amount of transcribed data from a challenging SLU task, to better semantic information extraction on this downstream task. Thus, we propose to enrich the SAMU-XLSR [Khurana et al., 2022] model with contextual information of dialog turns in addition to the previously acquired multilingual multimodal semantic information.

We are also interested in semantic information extraction from speech signals using end-to-end approaches. The performance of the Contextual-SAMU-XLSR model will be evaluated on SLU task in different languages and domains.

The experiments will be performed on two challenging SLU datasets. I) A new version of the MEDIA [Bonneau-Maynard et al., 2005] French corpus enriched with intent information in addition to the slots. II) The TARIC corpus [Masmoudi et al.,] in Tunisian dialect, enriched with semantic annotations (slots and dialog acts). Both corpora will be publicly available soon. In addition, we propose to use the DailyDialog [Li et al., 2017] corpus to enrich the SAMU-XLSR model with contextual information.

The objectives of the internship are :

- Extend the recent work [Laperrière et al., 2023] to develop an end-to-end SLU system for joint slot and intent detection on the new version of MEDIA TASK.
- Enrich the SAMU-XLSR model with contextual information of dialog turns
- Evaluate the performance of contextual SaMU XLSR representation on both corpora and investigate how the cross-lingual and cross-domain portability from distant languages could be beneficial to make the semantically enriched representation more accurate.

The SLU models will be implemented using the open-source SpeechBrain toolkit [?] dedicated to neural speech processing.

Expected profile

- Master 2 profile student in Computer Science, specialized at least in one of the following topics :
 - Machine learning
 - Natural language processing
- Technical skills : python, linux

Practical information

- Duration of internship : 5-6 months
- Beginning of the internship : start date is to be defined with the intern, but preferably January or February
- Gratification : around 660 /month and reimbursement of transport costs and canteen subsidy

Références

- [Bonneau-Maynard et al., 2005] Bonneau-Maynard, H., Rosset, S., Ayache, C., Kuhn, A., and Mostefa, D. (2005). Semantic annotation of the french media dialog corpus. In *Interspeech*.
- [Khurana et al., 2022] Khurana, S., Laurent, A., and Glass, J. (2022). Samu-xlsr : Semantically-aligned multi-modal utterance-level cross-lingual speech representation. *IEEE Journal of Selected Topics in Signal Processing*, 16(6) :1493–1504.
- [Laperrière et al., 2023] Laperrière, G., Nguyen, H., Ghannay, S., Jabaian, B., and Estève, Y. (2023). Specialized semantic enrichment of speech representations. In *2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)*, pages 1–5.
- [Li et al., 2017] Li, Y., Su, H., Shen, X., Li, W., Cao, Z., and Niu, S. (2017). Dailydialog : A manually labelled multi-turn dialogue dataset. *ArXiv*, abs/1710.03957.
- [Masmoudi et al.,] Masmoudi, A., Esteve, Y., Belguith, L. H., and Habash, N. A corpus and phonetic dictionary for tunisian arabic speech recognition.
- [Tomashenko et al., 2020] Tomashenko, N., Raymond, C., Caubrière, A., Mori, R. D., and Estève, Y. (2020). Dialogue history integration into end-to-end signal-to-concept spoken language understanding systems. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8509–8513.