Apprentissage multi-tâche pour le traitement de la parole et de la langue dans le cadre de conversations spontanées multi-locuteurs

L'équipe R&D (https://labs.linagora.com/) de la société **LINAGORA** (http://linagora.com) développe en open-source des outils d'assistance intelligente pour entreprises, y compris l'assistant vocal LinTO (https://linto.ai/), et LinSTT (https://github.com/linto-ai/linstt-engine), un outil de reconnaissance de la parole qui est capable de transcrire sous forme textuelle un signal vocal, ce qui nous permet de produire, de manière automatique, des transcriptions de réunion. Actuellement, nous travaillons sur un gestionnaire de conversation, Conversation Manager, une plateforme qui permettra à partir d'un enregistrement complet d'une réunion d'en déduire un résumé aussi pertinent que possible. L'idée est qu'un utilisateur du Conversation Manager va pouvoir d'abord visualiser, corriger et annoter une transcription proposée par notre système et ensuite exploiter le contenu de la transcription et ses annotations pour créer un résumé de manière semi-automatique.

Pour ce faire, il est impératif que la transcription proposée à l'utilisateur, avant l'étape de correction, soit aussi correcte et facile à visualiser que possible, ce qui peut être difficile pour les transcriptions de réunion où il y a plusieurs locuteurs et où les participants ont tendance à faire des interventions longues et mal structurées d'un point de vue grammatical. Pouvoir bien associer un tour de parole à son locuteur (**segmentation et regroupement en locuteurs,** ou *diarisation* en anglais) et ajouter les marques de **ponctuation** qui rendent le texte plus facile à lire sont très importants pour faire des transcriptions de haute qualité.

La *diarisation* et la ponctuation peuvent ensuite servir à améliorer les algorithmes de résumé automatique en aidant un système à découper le contenu d'une réunion en clauses individuelles --- appelés **segments discursifs.** Ces segments fournissent des unités sémantiques qui seront passées ensuite aux algorithmes de résumé qui jugeront quels segments sont plus centraux à la conversation et du coup, au résumé final.

Pour ce stage, le stagiaire étudiera les trois tâches - la *diarisation*, la ponctuation, et la segmentation discursive - en parallèle avec une approche d'apprentissage multi-tâche. L'entraînement du modèle sera fait sur des données de conversation transcrites soit en français, soit en anglais. Nous commencerons avec des modèles existants de ponctuation et segmentation qui se basent sur une architecture de transformer + bi-LSTM ainsi qu'un modèle de diarisation. La nouveauté de ce stage consistera dans (a) l'approche multi-tâche pour étudier ces trois sujets en parallèle et (b) l'usage des informations acoustiques des enregistrements de conversation et de réunion (alors que les modèles de base pour la ponctuation et la segmentation discursive sont entraînés exclusivement sur du texte).

L'encadrement du stage : Le stagiaire sera encadré par Samir Tanfous de LINAGORA, mais travaillera en collaboration avec Julie Hunter de LINAGORA et plusieurs membres du laboratoire IRIT, notamment Philippe Muller de l'équipe Melodi (NLP) et Thomas Pellegrini et Hervé Bredin de l'équipe Samova (Traitement de la parole).

Localisation : LINAGORA, soit à Paris, soit à Toulouse

Compétences clés recherchées :

- Étudiants de M2 ou d'école d'ingénieur en dernière année, en informatique et IA avec des compétences en machine learning

- De l’expérience en deep learning et PyTorch serait un plus
- De l'expérience en speech processing et/ou NLP serait un plus

Durée du stage : 5-6 mois, début du stage dès que possible.

Gratification : à définir selon l'expérience du candidat

Contact email : stanfous@linagora.com, jhunter@linagora.com

Références

Bredin, H., Laurent, A. (2021) End-To-End Speaker Segmentation for Overlap-Aware Resegmentation. Proc. Interspeech 2021, 3111-3115.

Muller, P., Braud, C., Morey, M. (2019) ToNy: Contextual embeddings for accurate multilingual discourse segmentation of full documents. Proceedings of the Workshop on Discourse Relation Parsing and Treebanking 2019, 115-124.