

Etat d'art de LIUM Spkdiazation

L'outil LIUM Spkdiazation est un logiciel de segmentation et de diarisation automatique du signal audio, développé par le Laboratoire d'Informatique de l'Université du Mans (LIUM). Il permet notamment la détection du nombre de locuteurs et la différenciation du genre.

Détection du nombre de locuteurs : La méthode de diarisation de LIUM Spkdiazation se compose de deux étapes : la segmentation et le regroupement. La segmentation consiste à découper le signal audio en segments de parole, qui sont détectés à l'aide d'un détecteur de parole basé sur une fenêtre glissante. Le regroupement consiste ensuite à regrouper les segments de parole en clusters correspondant à un même locuteur. Pour cela, LIUM Spkdiazation utilise un modèle de GMM pour chaque locuteur, qui est entraîné à partir des caractéristiques acoustiques des segments de parole, telles que la fréquence fondamentale et la modulation de fréquence. Le nombre de locuteurs est estimé en utilisant un modèle de Markov caché, qui prend en compte les transitions entre les clusters de locuteurs.

Différenciation du genre : LIUM Spkdiazation utilise un modèle de GMM supervisé pour différencier le genre des locuteurs. Ce modèle est entraîné à partir d'un ensemble de données d'apprentissage, qui est annoté avec le genre des locuteurs. Le modèle est alors utilisé pour prédire le genre des locuteurs dans les enregistrements audio. Les caractéristiques acoustiques utilisées pour l'entraînement du modèle incluent la fréquence fondamentale, la modulation de fréquence, la durée des segments de parole, la variance de la durée, la moyenne de la durée et la variation de l'énergie. L'approche de LIUM Spkdiazation pour la différenciation du genre a été évaluée à travers plusieurs tests locaux, qui ont montré que le système avait une précision élevée pour la reconnaissance du genre masculin, mais des performances légèrement inférieures pour le genre féminin.

En résumé, LIUM Spkdiazation utilise une approche basée sur les modèles statistiques, tels que les HMM et les GMM, pour segmenter, regrouper et diariser les signaux audios, ainsi que pour différencier le genre des locuteurs.

—

Recherche :

Kaldi : Kaldi est un outil open source de reconnaissance de la parole. Tout comme le LIUM, il est également développé en C++ et est conçu pour être modulaire et extensible. Cependant, Kaldi a une portée plus large que le LIUM et prend en charge des tâches telles que la reconnaissance de la parole, la traduction automatique, la synthèse de la parole et la segmentation de signal audio.

Stanford CoreNLP : Stanford CoreNLP est une suite d'outils open source pour le traitement automatique des langues naturelles. Il prend en charge des tâches telles que l'analyse syntaxique, l'analyse sémantique, la reconnaissance d'entités nommées, la relation de co-référence et la dépendance de l'arbre. Comparé au LIUM, Stanford CoreNLP est plus largement utilisé et prend en charge plusieurs langues différentes.

Spacy : Spacy est une bibliothèque open source pour le traitement automatique des langues naturelles. Il prend en charge des tâches telles que l'analyse syntaxique, l'analyse sémantique, la reconnaissance d'entités nommées, la dépendance de l'arbre et la reconnaissance de relations. Comparé au LIUM, Spacy est plus facile à utiliser et possède une communauté plus active.

Google Cloud Natural Language : Google Cloud Natural Language est un outil cloud pour le traitement automatique des langues naturelles. Il prend en charge des tâches telles que l'analyse syntaxique, la reconnaissance d'entités nommées, l'analyse de sentiments et la classification de texte. Comparé au LIUM, Google Cloud Natural Language est plus facile à utiliser et est conçu pour être utilisé dans des applications d'entreprise.

—

Définitions :

- **Diarisation** : La diarisation est le processus de segmentation d'un enregistrement audio en segments de parole, et de regroupement de ces segments en clusters correspondant à un même locuteur. La diarisation est souvent utilisée pour l'analyse de conversations enregistrées, la reconnaissance de la parole, ou la création de corpus de données pour l'apprentissage automatique.
- **Modèle de Markov caché (HMM)** : Un modèle de Markov caché est un modèle probabiliste utilisé pour modéliser les séquences de données, telles que les signaux audios. Le modèle HMM consiste en un ensemble d'états cachés qui évoluent au fil du temps selon une matrice de transition probabiliste, et émettent des symboles observables selon des distributions de probabilité spécifiques à chaque état caché. Les HMM sont souvent utilisés pour la reconnaissance de la parole, la segmentation de séquences, ou la classification de données temporelles.
- **Classification de mélange gaussien (GMM)** : La classification de mélange gaussien est une méthode de classification supervisée utilisée pour modéliser la distribution de probabilité des caractéristiques d'un ensemble de données. La méthode de classification de mélange gaussien suppose que les données proviennent de plusieurs distributions gaussiennes, et utilise un algorithme d'apprentissage pour estimer les paramètres de ces distributions. Les GMM sont souvent utilisés pour la reconnaissance de la parole, la diarisation, ou la classification de données en général.
- **Caractéristiques acoustiques** : Les caractéristiques acoustiques sont des mesures des propriétés acoustiques d'un signal audio, telles que la fréquence fondamentale, la modulation de fréquence, la durée, ou l'énergie. Les caractéristiques acoustiques sont souvent utilisées pour l'analyse de la parole, la reconnaissance de la parole, ou la diarisation.
- **Ensemble de données d'apprentissage** : L'ensemble de données d'apprentissage est un ensemble de données utilisé pour entraîner un modèle d'apprentissage automatique. L'ensemble de données d'apprentissage est souvent annoté avec des

étiquettes qui indiquent les classes des données, telles que les locuteurs ou les genres. Les ensembles de données d'apprentissage sont souvent utilisés pour l'apprentissage supervisé des modèles d'apprentissage automatique, tels que les modèles de GMM ou les réseaux de neurones.