

Master Informatique

Parcours Données et Connaissances

Projet Inter-Promo 2020 : Analyse de données multimodales pour la caractérisation du déroulement de réunions

Isabelle Ferrané - Lionel Pibre - Véronique Moriceau
Isabelle.Ferrane@irit.fr, Lionel.Pibre@irit.fr
Veronique.Moriceau@irit.fr

Groupes : 1 groupe M2

Mots clés : Multimodalité, reconnaissances d'objets, segmentation audio, analyse thématique, actes de dialogues, machine learning

1 Contexte du projet

Ce projet se situe dans le contexte de la compréhension automatique de contenus multimodaux. Il porte donc sur le traitement de données multimodales (audio, vidéo et texte) pour caractériser le contenu d'un ensemble de vidéos et permettre l'accès à des informations complexes comme par exemple la situation d'interaction, le rôle des participants, la structuration en séquences thématiques, etc.

Le projet s'inscrit dans le cadre du projet LinTo¹, dont le but est de développer un assistant conversationnel intelligent, open-source, dédié aux entreprises.

2 Corpus, Modalités, Outils de prétraitement à disposition

2.1 Corpus audio-vidéo AMI

L'assistant dans le projet LinTo, devra communiquer en français. Les corpus en français de retranscription étant en cours d'élaboration, le projet se construira sur des données en anglais, issus du projet AMI².

Le corpus dans AMI contient 100 heures de réunion (en anglais) enregistrées en utilisant plusieurs appareils d'enregistrement synchronisés (vidéos, micros, tableaux électroniques, etc.), pour un total de 171 réunions. Les réunions sont de deux types : (1) les réunions à base de scénario (*scenario-based meetings*) regroupent 3 à 4 réunions simulant les réunions de quatre participants devant concevoir un système, et (2) les réunions sans scénarios qui sont de vraies réunions qui ont été enregistrées. Le corpus AMI a été transcrit (transcription orthographique) et annoté pour couvrir divers phénomènes (actes de dialogues, mouvement de la tête, etc.). Les participants à ces

1. <https://www.irit.fr/projet-pia-linto/>

2. <http://groups.inf.ed.ac.uk/ami/corpus/>

réunions sont principalement des anglophones non-natifs.

2.2 Modalités

Nous nous intéresserons dans un premier temps à une sous-partie du corpus (scenario-based meetings) et aux modalités suivantes :

- **Modalité VIDÉO** : enregistrements réalisés par une caméra donnant un point de vue général sur la pièce où se tient la réunion et par les caméras des ordinateurs personnels des participants. Dans ce second cas, chaque caméra filme une personne en gros plan.
- **Modalité AUDIO** : les conversations sont enregistrées par un micro d’ambiance ou via le microphone de l’ordinateur ou via le micro-casque de chaque participant.
- **Modalité TEXTE** : transcription manuelle des propos tenus (en anglais) par chaque participant et annotation fines associées.

Plus de détails sur les données et les annotations disponibles sont donnés dans le document correspondant au schéma d’annotation³. Le corpus est une ressource publique disponible qui peut être téléchargé⁴.

2.3 Outils de prétraitement

De nombreux outils automatiques sont disponibles et permette d’effectuer des premiers traitements aussi bien sur la modalité audio (détection d’activité vocale, segmentation et regroupement en locuteurs, qualification de l’environnement sonore bruité ou pas) que vidéo (localisation des personnes, détection des visages, détection d’objets, ...) ou texte (fréquence des mots, structure syntaxique des phrases, représentation vectorielle des mots ou des phrases, extraction d’entités nommées, ...). Une liste non exhaustive est donnée dans la table 1

- **Outils d’analyse VIDÉO** : des outils de détection et reconnaissance d’objets, de détection de visages pourront être utilisés.
- **Outils d’analyse AUDIO** : le flux audio pourra être segmenté en zone de parole/non parole, en locuteur, ...
- **Outils d’analyse TAL** : la transcription manuelle pourra être analysée d’un point de vue lexical, syntaxique, sémantique, dialogique.

3 Tâches à réaliser

3.1 Première phase : prise en main des outils et du corpus

Avant de traiter les données à disposition avec les outils proposés, il est nécessaire de :

- récupérer et d’étudier le contenu du corpus ;
- comprendre comment il est structuré et quelles sont les annotations disponibles ;
- récupérer et installer les outils de traitement des différentes modalités ;
- vérifier le bon fonctionnement de ces outils (des données de tests sont parfois fournies avec les outils pour s’assurer que le comportement de l’outil est correct) ;

3. <http://groups.inf.ed.ac.uk/ami/corpus/annotation.shtml>

4. <http://groups.inf.ed.ac.uk/ami/download/>

Outils	Fonctionnalité	Lien
OPEN CV	Face detection	https://www.pyimagesearch.com/
YOLO	Object Detection	https://pjreddie.com/darknet/yolo/
WebRTC	Voice Activity detection	https://github.com/wiseman/py-webrtcvad
PyBK	Speaker Diarization	https://github.com/josepatino/pyBK
Spacy	TAL and Named Entities extraction	https://spacy.io/
Autres outils	vus en cours ou trouvés dans vos recherches	

TABLE 1 – Liste des outils possibles

- de tester les outils sur les fichiers du corpus (prendre les données correspondant à une même réunion pour pouvoir vérifier la cohérence des résultats);
- de comparer les performances d’outils qui pourraient faire des traitements équivalents pour savoir lequel choisir en fonction du contexte;
- organiser les données pour disposer d’une partie apprentissage, développement et test qui seront utiles suivant la nature des méthodes utilisées.

Livrable Vous rendrez un premier document qui permettra de synthétiser les différentes étapes réalisées dans cette première phase.

3.2 Seconde phase : approche incrémentale

Après vous être intéressés à chaque modalité prise séparément afin d’en explorer les possibilités, vous pourrez ensuite procéder par fusion de modalités, c’est-à-dire étudier les combinaisons possibles pour rendre les caractérisations plus robustes et aussi plus riches.

Exemple de combinaison possibles Par exemple, associer temporellement des informations audio et visuelle comme la voix et le visage permet de caractériser chaque participant à la réunion. Associer des informations audio et textuelle comme la personne qui parle (tour de parole d’un locuteur) et ce qui est dit (entités nommées ou actes de dialogue) permet de déterminer comment la parole se distribue (qui dit quoi, participant très actif ou peu actif) voire d’identifier le rôle des participants (animateur de la réunion, ...) ou de typer les séquences (début de réunion, exposé, questions/réponses, débat, fin de réunion...) etc.

Etat de l’art sur les méthodes de fusion Une étude sur les méthodes de fusion pourra être entreprise pour avoir une idée des approches possibles et implémentable dans le cadre de ce projet.

Objectifs Il s’agit de réfléchir à ce qu’il est possible de faire compte tenu de vos connaissances (expérience acquise sur la phase 1 et enseignements spécifiques de la formation sur la partie TAL), de choisir un ou deux axes de travail et de les prioriser. L’objectif final du projet est de pouvoir traiter chaque réunion et d’extraire des informations sur sa structure et sur le déroulement de la réunion.

Livrables Vous rendrez un second document contenant un petit état de l'art et décrivant les développements réalisés ainsi que les résultats obtenus. L'évaluation du travail portera non seulement sur les résultats mais également sur la démarche scientifique adoptée.

3.3 Justification des choix

Une liste non exhaustive des outils possibles est donnée dans la table 1. Vous pouvez également intégrer d'autres outils dont vous auriez connaissance et qui aideraient à la caractérisation de ces contenus. La justification des choix est un aspect important. Dire, j'ai utilisé cet outil parce que je le connais n'est pas un bon argument si ces performances sont moins bonnes qu'un autre outil que vous ne connaissez pas". Des liens vers des publications complémentaires vous seront données en début de projet.

4 Logiciels/outils

Les développements se feront en python. Les différents outils et bibliothèques mentionnés plus haut sont utilisables dans un environnement python. Vous pouvez utiliser google colab, conda, jupyter pour réaliser vos traitements et garder une trace du déroulement de vos expérimentations (notebooks à rendre).