

Comment les IA peuvent-elles apprendre la langue des signes ?

Hannah Bull

Les langues des signes sont un moyen de communication essentiel pour les communautés sourdes. Elles sont des langues visuo-gestuelles, qui utilisent comme modalités les mains, les expressions faciales, le regard et les mouvements du corps. Elles ont des structures grammaticales complexes et des lexiques riches qui sont considérablement différents de ceux que l'on trouve dans les langues parlées. Les spécificités des langues des signes en termes de canaux de communication, de structure et de grammaire rend le développement des technologies d'IA pour la langue des signes très difficile.

Les performances des systèmes de traduction automatique par l'IA entre des langues écrites ou parlées sont actuellement suffisantes pour de nombreux cas d'utilisation quotidienne, tels que la traduction de vidéos, de sites web, d'e-mails et de documents. En revanche, les systèmes de traduction automatique par l'IA pour les langues des signes n'existent pas en dehors de cas d'utilisation très spécifiques avec un vocabulaire limité. La traduction automatique de langues des signes est un défi pour deux raisons principales. Premièrement, les langues des signes sont des langues à faibles ressources avec peu de données d'entraînement disponibles. Deuxièmement, les langues des signes sont des langues visuelles et spatiales sans forme écrite, naturellement représentées sous forme de vidéo plutôt que d'audio ou de texte.

Pour relever le premier défi, nous fournissons de grands corpus de données pour l'entraînement et l'évaluation des systèmes de traduction automatique en langue des signes, avec des contenus vidéo en langue des signes interprétée et originale, ainsi que des sous-titres écrits. Alors que les données interprétées nous permettent de collecter un grand nombre d'heures de vidéos, les vidéos originalement en langue des signes sont plus représentatives de l'utilisation de la langue des signes au sein des communautés sourdes. Les sous-titres écrits peuvent être utilisés pour apprendre automatiquement différents aspects de la langue des signes.

Pour relever le deuxième défi, ce projet propose des méthodes permettant de mieux comprendre les vidéos en langue des signes. Alors que la segmentation des phrases est généralement triviale pour les langues écrites grâce à la ponctuation, la segmentation des vidéos en langue des signes en phrases repose sur la détection

d'indices sémantiques et prosodiques subtils dans les vidéos. Nous utilisons des indices prosodiques pour apprendre à segmenter automatiquement une vidéo en langue des signes en unités de type phrase, déterminées par les limites des sous-titres. En développant cette méthode de segmentation, nous apprenons ensuite à aligner les sous-titres du texte sur les segments de la vidéo en langue des signes en utilisant des indices sémantiques et prosodiques, afin de créer des paires au niveau de la phrase entre la vidéo en langue des signes et le texte. Cette tâche est particulièrement importante pour les données interprétées, où les sous-titres sont généralement alignés sur l'audio et non sur la langue des signes. En utilisant ces paires vidéo-texte alignées automatiquement, nous développons et améliorons plusieurs méthodes différentes pour reconnaître de très nombreux signes lexicaux, en interrogeant des mots dans le texte des sous-titres et en recherchant des indices visuels dans la vidéo en langue des signes pour les signes correspondants.

Les résultats de ce projet sont les suivants :

1. un corpus de la langue des signes française avec des traductions écrites en français : un corpus rare avec contenu original en langue des signes produit en dehors de contexte d'un laboratoire
2. un corpus de vidéos interprétées en langue des signes britannique avec des sous-titres anglais : le plus grand corpus actuellement disponible pour la recherche en langue des signes
3. des résultats de référence pour la nouvelle tâche de segmentation en phrases des vidéos en langue des signes
4. une méthode de pointe pour aligner automatiquement les sous-titres à la langue des signes
5. des méthodes améliorées et de nouvelles méthodes pour reconnaître de très nombreux signes dans des vidéos