

"Analyse : que retirons-nous de ces expériences ?"

Subjectivité et objectivité des mesures dans l'évaluation des systèmes interactifs humanoïdes : corrélations et limites

Magalie Ochs – Informatique

Aujourd'hui, les systèmes interactifs humanoïdes (personnages virtuels, avatars ou robots humanoïdes) sont principalement évalués à travers des questionnaires subjectifs post-expérience. Ces questionnaires permettent à la fois de mesurer la perception de l'utilisateur mais aussi son expérience. Plusieurs questionnaires ont été proposés et sont largement utilisés par exemple pour mesurer la crédibilité du système, les émotions de l'utilisateur, le sentiment de présence et de co-présence ou encore la confiance. Certaines dimensions peuvent être partiellement mesurées à travers des indices plus objectifs (e.g. les expressions faciales pour les émotions, la conductivité de la peau pour le stress ou encore le temps de réponse pour la difficulté de la tâche). Ces indices présentent cependant un certain nombre de limites qui peuvent être dues à un environnement bruité ou au caractère intrusif des capteurs. De plus, il existe aujourd'hui un très grand nombre de dimensions subjectives pour lesquelles des indices objectifs n'ont pas été identifiés, par exemple l'engagement ou encore la confiance. Un des enjeux dans l'évaluation des systèmes interactifs est de pouvoir identifier ces mesures objectives. Aujourd'hui, les algorithmes d'apprentissage automatique nous permettent *a priori* de faire le lien entre des indices comportementaux enregistrés pendant une évaluation et des mesures subjectives collectées à travers un questionnaire. À partir des données d'une expérimentation, il serait alors possible d'identifier les indices comportementaux objectifs caractéristiques des mesures subjectives. Plusieurs travaux de recherche ont été menés dans ce sens. Par exemple, dans (Ochs et al., 2022), nous avons collecté un corpus d'interactions humain-agent virtuel lors d'une évaluation d'une patiente virtuelle utilisée pour l'entraînement des médecins à l'annonce d'événements indésirables graves. À la fois le comportement multimodal des utilisateurs (mouvements du corps et voix) et leurs perceptions subjectives de l'agent ont été enregistrés. À partir de ces données, nous avons exploré les corrélations entre le comportement multimodal de l'utilisateur et son sentiment subjectif en utilisant des algorithmes d'apprentissage automatique. En particulier, nous nous sommes intéressés aux indices comportementaux de l'engagement de l'utilisateur qui se traduit en réalité virtuelle par le sentiment de présence et de co-présence. Les performances des modèles d'apprentissage montrent qu'effectivement le niveau d'engagement de l'utilisateur peut être efficacement prédit à partir d'indices multimodaux comme les mouvements de la tête ou la longueur des phrases. Ces résultats démontrent donc l'existence d'indices comportementaux reflétant l'engagement de l'utilisateur. Dans une autre expérience, nous avons comparé le comportement d'un utilisateur suivant qu'elle/il interagissait avec un humain, un robot ou un agent virtuel (Hmamouche et al., 2022). Notre objectif était de pouvoir identifier s'il existait des indices objectifs dans le comportement verbal de l'humain permettant de refléter une interaction naturelle (i.e. comme avec un autre humain). En utilisant des algorithmes d'apprentissage profond, nous avons pu démontrer qu'effectivement certaines caractéristiques vocales étaient particulièrement importantes pour pouvoir prédire une interaction naturelle, comme par exemple la richesse lexicale ou encore la polarité des mots employés.

Cette problématique de recherche visant à une évaluation objective des systèmes interactifs pose encore plusieurs questions. Tout d'abord, est-ce possible d'objectiver d'un point de vue comportemental un sentiment subjectif ? d'un point de vue apprentissage automatique, quelle quantité de données est nécessaire ? comment rendre interprétable les modèles en définissant des caractéristiques hauts-niveaux explicables ? Quels modèles sont le plus adaptés pour l'explicabilité ? existe-t-il un modèle générique pour certains sentiments subjectifs ? Sinon, comment prendre en compte les variations interindividuelles ? existe-il des verrous sur les briques de reconnaissance des indices comportementaux à considérer dans la construction des modèles ?

Références

Ochs, M., Bousquet, J., Pergandi, J. M., & Blache, P. (2022). Multimodal behavioral cues analysis of the sense of presence and co-presence during a social interaction with a virtual patient. *Frontiers in Computer Science*.

Hmamouche, Y., Ochs, M., Chaminade, T., & Prévot, L. (2022, August). Listen and tell me who the user is talking to: Automatic detection of the interlocutor's type during a conversation. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)* (pp. 660-665). IEEE.