Interactions naturelles en temps réel dans les communautés mixtes humains - compagnons artificiels

Mots-clefs : Interactions humains-agents, Interactions multimodales, Interactions multi-parties, agents conversationnels animés, robots

Encadrement : Julien Saunier, Alexandre Pauchet

Localisation : LITIS, INSA Rouen Normandie (Rouen, France)

Date prévue de début de thèse : Octobre 2023

Rémunération : environ 1 650 euros / mois

Candidature : CV, notes de M1 et M2, lettre de motivation et lettres de recommandation sont à faire parvenir à alexandre.pauchet@insa-rouen.fr et julien.saunier@insa-rouen.fr.

Contexte :
La croissance du nombre d'objets connectés, des robots assistants et des interfaces humain-machines intuitives et naturelles a permis une démocratisation croissante des systèmes cyber-physiques et socio-techniques. Il s'agit de systèmes intégrant à la fois des utilisateurs humains, des robots et des agents artificiels, en interactions sociales dans des environnements réels, augmentés ou virtuels.

Si les interactions avec un seul utilisateur en tour par tour disposent d'une littérature abondante, tous les verrous scientifiques et techniques ne sont pas encore levés concernant les interactions dialogiques. La mise en place de systèmes coopératifs intégrant plusieurs utilisateurs humains, agents virtuels et robots reste difficile. Depuis l'avènement de l'informatique affective (Picard, 1997), des Agents Conversationnels Animés -ACA- (Cassel, 2000), et des robots sociaux (Gockley et al., 2005), qui mettent en œuvre des systèmes prenant en compte les émotions des utilisateurs et pouvant en « jouer », n'est que très peu traitée dans un contexte multi-partie. En particulier, la capacité d'un agent collaboratif à répondre « just in time » à son contexte est peu développée : la plupart des systèmes attendent la fin de chaque tour de parole avant d'interpréter les énoncés des utilisateurs et de décider de la prochaine action communicative à effectuer. Ils ont ainsi un temps de réaction bien supérieur à celui d'un humain, introduisant des ruptures défavorables à l'interaction.

Objectifs scientifiques
L'objectif de cette thèse est d'étudier les moyens théoriques et pratiques pour permettre une interaction multimodale et multi-parties en temps quasi réel afin d'en améliorer la fluidité et l'acceptabilité. Ces travaux comportent deux originalités principales : (1) la gestion au fur et à mesure du dialogue multimodal aussi bien en entrée qu'en sortie, permettant aux compagnons virtuels et robotiques d'améliorer leurs capacités communicatives, verbales ou non-verbales, et (2) la dimension multi-parties, c'est-à-dire intégrant plusieurs agents virtuels, robots et humains. Il s'agit de concevoir des modèles et protocoles d'interaction multi-parties, que cette interaction se déroule uniquement entre agents autonomes aussi bien qu'entre agents autonomes et humains dans le cadre de sociétés mixtes. Une façon de générer ces modèles d'interaction est d'utiliser des mécanismes d'apprentissage automatique.

Une des difficultés consistera donc à intégrer des mécanismes dits « just in time » dans un contexte multimodal et multi-parties, permettant la prise en compte de tous les membres d'équipes mixtes agents-humains, dans lesquels les effectifs (nombre d'agents, nombre d'humains dans l'équipe) peuvent varier.

Nous nous focaliserons sur l'utilisation d'ACA et de robots. Dans ce contexte, la maturation des outils pour la réalité mixte (réalité virtuelle et réalité augmentée), que ce soit en termes logiciel (e.g. Unity3D, Unreal Engine, ARToolKit, ...) comme matériel accessible au grand public (Oculus Rift, HTC Vive, Playstation VR, …), et le développement des robots sociaux (Nao, Pepper, …) permet d'envisager de nouvelles façons d'intégrer ces agents au sein d'environnements mixtes réel/virtuel.

Projet détaillé

Les systèmes cyber-physiques et socio-techniques, constitués à la fois d'utilisateurs humains, de robots et d'agents artificiels en interactions sociales, se démocratisent. Si cette démocratisation est un élément majeur pour proposer de nouveaux services au quotidien, sa diffusion se heurte à deux verrous principaux : d'une part, la reconnaissance de l'activité humaine reste imprécise, tant au niveau opérationnel (localisation, cartographie, identification d'objets et d'utilisateurs) que cognitif (reconnaissance et suivi d'intention). D'autre part, l'interaction passe par des vecteurs différents qu'il faut adapter en fonction du contexte (robotique, réalité mixte, réalité virtuelle), de l'utilisateur et de la situation ou tâche en cours.

(Cassell, 2000) a introduit le concept d'interaction sociale en face-à-face avec un agent animé. Le niveaux de détails utilisés pour représenter un personnage virtuel souvent très élevés induisent des attentes particulières de la part de l'interlocuteur quant aux capacités interactives de l'agent. Cependant, les faibles compétences conversationnelles et les réactions non naturelles des agents, qu'ils soient virtuels ou robotiques, déçoivent et mènent à des interactions peu naturelles entre l'humain et le système. Ce phénomène est appelé la "vallée dérangeante" (the uncanny valley) (Mori, 1970). Ce type de comportements affectent l'engagement des utilisateurs envers les agents (Beale et al, 2009). Pour surmonter ces inconvénients, l'agent doit répondre à la frustration des utilisateurs (Klein et al, 2002), devenir plus empathique (Ochs et al, 2008 ; Prendinger et al, 2005), émotionnel (Poggi et al, 2005) et réagir au moment approprié avec une posture ou un geste adapté à la situation (Prepin et al, 2013). Des études pédagogiques (Moundrido et al, 2002) et utilisant les jeux sérieux (Prendinger et al., 2003) ont montré l'existence d'un lien entre la présence d'un personnage virtuel et les performances d'apprentissage et d'engagement de l'utilisateur. Des études similaires existent également avec des robots. Le même niveau d'engagement est observé chez les enfants (Kozima et al, 2005 ; Robins et al, 2005), en situation de tutorat (Han et al., 2005) ou dans des situations de développement cognitif précoce (Von Hofsten et al, 2007).

Dans ce cadre, l'objectif de cette thèse est ainsi d'explorer les possibilités d'exploitation de modèles d'interaction multimodale et multi-parties en temps quasi réel, afin d'améliorer l'acceptabilité de ces systèmes interactifs.

Le premier verrou à lever concerne la conception d'un mécanisme permettant la prise en compte de tous les membres de communautés mixtes agents-humains, dans lesquels les effectifs (nombre d'agents, nombre d'humains) et les représentations (agents virtuels ou robots d'une part, humain en personne ou avatar d'autre part) peuvent varier. Ces mécanismes se retrouvent aussi bien au niveau des agents que du système (environnement, plateforme) (Weyns et al, 2015 ; Rincon et al, 2016). La modélisation de systèmes intégrant humains et agents dans des environnements réels et virtuels, c'est-à-dire de systèmes qui doivent mêler au sein d'un même espace une couche physique et une couche logique/virtuelle pour permettre l'interaction, pourra s'appuyer sur la notion de cognition incarnée. L'objectif est de considérer de façon transparente et unifiée utilisateurs, robots et agents virtuels à l'aide d'une interface commune.

L'étudiant recruté en thèse se focalisera également sur la modélisation de la composante affective pour l'étude de la contagion émotionnelle (Barsade, 2002) dans le contexte d'une tâche de travail collaboratif. L'objectif est ainsi de rendre les agents, qu'il s'agisse d'agents virtuels ou de robots, autonomes dans leurs choix des actions dialogiques notamment à composante émotionnelle. Leurs comportements dépendent alors à la fois du dialogue et de l'état inféré des autres participants de l'interaction (humains ou agents). L'intégration de cette composante émotionnelle permet d'augmenter la crédibilités des agents ainsi que l'engagement dans l'interaction des utilisateurs (Picard, 1997 ; Cassel, 2000).

Le troisième aspect concerne la conception et l'apprentissage de protocoles d'interaction pour lagestion de dialogues affectifs, multi-modaux et multi-parties. Pour cela, une première approche concerne la mise en place d'un système permettant au concepteur d'agents de se focaliser sur les choix dialogiques, orientés tâche. Concernant la gestion du dialogue, une approche exploitant nos travaux passés par extraction de motifs dialogiques (Alès, 2018) et l'apprentissage de modèles d'interaction multimodale et multi-parties (Malik, 2021) sera privilégiée.

Le dernier verrou concerne l'intégration des aspect temps quasi-réel à ces travaux. Il nécessitera

- de considérer des modèles capables de traiter un flux multimodal, afin de capturer, représenter et interpréter à tout moment le comportement et les intentions des utilisateurs impliqués dans l'interaction ;
- de sélectionner en contexte incertain un comportement communicatif pertinent ;
- de pouvoir interrompre ou modifier instantanément et de manière cohérente et fluide un comportement interactif en cours d'exécution ;
- de réaliser en parallèle, de manière hiérarchique, itérative et interruptible, les étapes de reconnaissance des comportements utilisateur et les actions communicatives de l'agent.

Ainsi, il s'agira à la fois de se concentrer sur la captation de l'état de l'utilisateur (Rasendrasoa, 2022), de la gestion des tours de parole (Jégou, 2018 ; Malik, 2020) mais également de définir une architecture d'agent incrémentale (Kopp, 2014), multimodale et empathique (Barange, 2022) permettant de jouer des comportements communicatifs adaptés à la situation dialogique.

Plan de travail proposé

1. État de l'art
    1. Dialogue humain-agent/robot, dialogue multimodale, dialogue multi-partie
    2. Informatique affective, attitudes émotionnelles
    3. Architectures agents et SMA
    4. Apprentissage incrémental de modèle d'interaction
2. Modélisation
    1. Définition d'une architecture incrémentation pour l'interaction
    2. Modèles multimodal, multi-parties, en temps quasi-réel
    3. Gestion des tours de parole
3. Implémentation
4. Évaluation
    1. Mise au point du scénario et des outils de validation
    2. Intégration du scénario aux différentes plateformes
    3. Expérimentations
5. Validation

Références bibliographiques

1. Alès Z., Pauchet A., Knippel A.: Extraction and Clustering of Two-Dimensional Dialogue Patterns. Int. J. Artif. Intell. Tools 27(2): 1850001:1-1850001:27 (2018)
2. Barange M., Rasendrasoa S., Bouabdelli M., Saunier J., Pauchet A.: Multimodal adaptive empathic agent architecture. IVA 2022: 36:1-36:3
3. Barsade, S. G. (2002). The ripple effect: Emotional contagion and its influence on group behavior. Administrative Science Quarterly, 47(4), 644-675.
4. Beale R., Creed C. (2009). Affective interaction: How emotional agents affect users. International Journal of Human-Computer Studies, vol. 67, n o 9, p. 755–776.
5. Cassell, J., Nudge nudge wink wink: Elements of face-to-face conversation for embodied conversational agents, Embodied conversational agents (2000) 1–27.
6. R. Gockley, A. Bruce, J. Forlizzi, M. Michalowski, A. Mundell, S. Rosenthal, B. Sellner, R. Simmons, K. Snipes, A.C. Schultz, Jue Wang : "Designing robots for long-term social interaction", IEEE/RSJ International Conference on Intelligent Robots and Systems, Edmonton, Alta., pp. 1338-1343, 2005.
7. Jégou, M., Chevaillier, P., A computational model for the emergence of turn-taking behaviors in user-agent interactions. Journal of Multimodal User Interfaces 12(3), pp. 199-223, 2018.
8. Kopp, S., van Welberger, H., Yaghoubzadeh Torky, R., Buschmeier, H., An Architecture for Fluid Real-time Conversational Agents: Integrating Incremental Output Generation and Input Processing, Journal on Multimodal User Interfaces 8(1), pp. 97–108, 2014.
9. Kozima H., Nakagawa C., Yasuda Y. (2005). Interactive robots for communication-care: a case-study in autism therapy. In Robot and human interactive communication, p. 341-346.
10. Klein J., Moon Y., Picard R. W. (2002). This computer responds to user frustration: Theory, design, and results. Interacting with computers, vol. 14, n o 2, p. 119–140.

11. Malik U., Saunier J., Funakoshi K., Pauchet A ;: Who Speaks Next? Turn Change and Next Speaker Prediction in Multimodal Multiparty Interaction. ICTAI 2020: 349-354.
12. Malik M., Barange M., Saunier J., Pauchet A.: A novel focus encoding scheme for addressee detection in multiparty interaction using machine learning algorithms. J. Multimodal User Interfaces 15(2): 1-14 (2021). Moundridou M., Virvou M. (2002). Evaluating the persona effect of an interface agent in a tutoring system. Journal of computer assisted learning, vol. 18, n°3, p. 253–261.
13. Ochs M., Pelachaud C., Sadek D. (2008). An empathic virtual dialog agent to improve human-machine interaction. In Proceedings of the 7th international joint conference on autono mous agents and multiagent systems-volume 1, p. 89–96.
14. Picard, R. W. (1997). Affective computing (Vol. 252). Cambridge: MIT press.
15. Poggi I., Pelachaud C., Rosis F., Carofiglio V., Carolis B. (2005). Greta. a believable embodied conversational agent. Multimodal intelligent information presentation, p. 3–25.
16. Prendinger H., Mayer S., Mori J., Ishizuka M. (2003). Persona effect revisited. In Intelligent virtual agents, p. 283–291. Berlin Heidelberg.
17. Prendinger H., Ishizuka M. (2005). The empathic companion: A character-based interface that addresses users' affective states. Applied Artificial Intelligence, vol. 19, n o 3-4, p. 267–285.
18. Prepin K., Pelachaud C. (2013). Basics of intersubjectivity dynamics: Model of synchrony emergence when dialogue partners understand each other. In Agents and artificial intelligence, p. 302–318. Springer.
19. Rasendrasoa S., Pauchet A., Saunier J., Adam S.: Real-Time Multimodal Emotion Recognition in Conversation for Multi-Party Interactions. ICMI 2022: 395-403Rincon, J. A., Julian, V., & Carrascosa, C. (2016). Developing an emotional-based application for human-agent societies. Soft Computing, 20(11), 4217-4228.
20. Von Hofsten C., Rosander K. (2007). From action to cognition (vol. 164). Elsevier Science.
21. Weyns, D., F. Michel, H. Van Dyke Parunak, O. Boissier, M. Schumacher, A. Ricci, A. Brandao, C. Carrascosa, O. Dikenelli, S. Galland, A. Pijoan, P. Simo Kanmeugne, J. A. Rodriguez-Aguilar, J. Saunier, V. Urovi and Franco Zambonelli (2015) : Agent Environments for Multi-Agent Systems -- A Research Roadmap. In Agent Environments for Multi-Agent Systems IV (pp. 3-21). Springer International Publishing.