

Interactions Homme-Machine Évoluées

Synthèses bibliographiques et Projets

ASI5 - INSA Rouen

1 Travail à réaliser

L'objectif de ce travail est tout d'abord de réaliser une synthèse bibliographique présentant l'état des connaissances sur un thème donné, pour lequel les solutions technologiques ne sont pas forcément encore complètement développées au niveau industriel. Il vous faudra donc explorer un ensemble d'articles scientifiques (issus aussi bien du monde académique que de services R&D), afin d'évaluer l'état actuel de la recherche dans ce domaine, d'en effectuer une synthèse et de proposer une solution au problème proposé.

Dans un second temps (en IHME mais pas en IDA), les connaissances acquises seront utilisées sur la réalisation d'un projet. Les sujets de synthèse bibliographique proposés sont donc en rapport avec les projets.

N'hésitez pas à proposer votre propre sujet (de projet et/ou de synthèse bibliographique), à partir du moment où celui-ci est en relation avec les thématiques du cours.

1.1 Synthèses bibliographiques

Ce travail est à réaliser seul, binôme ou trinôme. Plusieurs groupes peuvent travailler sur le même sujet.

Le résultat de votre recherche se présentera sous la forme d'un rapport d'une longueur comprise entre 2 et 3 pages (en simple ou double colonne), hors liste de références bibliographiques. Ce rapport décrira votre travail d'analyse, de comparaison, voire de recommandation, sur une thématique donnée.

Afin de faciliter et amorcer votre recherche, 2 références initiales vous sont fournies ; elles vous permettront également de bien appréhender le travail qui vous est demandé car elles contiennent toujours une section concernant les travaux existants. Si vous en ressentez le besoin, un cours sur les synthèses bibliographiques est disponible sur Moodle (https://moodle.insa-rouen.fr/pluginfile.php/80454/mod_resource/content/1/SyntheseBibliographique.pdf). Une synthèse bibliographique de bonne facture référence en général entre 10 et 30 articles scientifiques.

1.2 Projets (IHME uniquement)

Cette partie du travail est à réaliser par groupes de 3 (et un groupe de 4) personnes (les groupes peuvent être différents de ceux des synthèses bibliographiques). Votre projet sera rendu sous la forme d'une archive contenant :

- votre code (lien gitlab par exemple) ;
- un rapport présentant votre travail (conception et implémentation) ;
- des manuels d'installation et utilisateur ;
- une vidéo de démonstration.

Même si ce n'est pas une obligation, il est conseillé de réaliser un projet en relation avec une des synthèses bibliographiques associées et réciproquement. Il vous est également possible d'avoir accès à l'ensemble des synthèses bibliographiques en relation avec votre sujet de projet.

Les sujets proposés peuvent être solutionnés avec n'importe quelle approche, qu'elle contienne une dimension apprentissage / science des données ou non, à partir du moment où celle-ci est correctement justifiée (par exemple par la synthèse bibliographique). Le panel des sujets proposés étant large, il vous est demandé dans l'introduction de vos rapports de positionner correctement le problème attaqué. Par ailleurs, il vous est tout à fait possible d'y ajouter une application pratique ou ludique afin d'illustrer et exploiter votre travail. Cela peut également être fait sous la forme d'un projet découpé entre plusieurs groupes (par exemple intégration de la reconnaissance d'expressions faciales - Gr1 - à un jeu de Poker avec des IA - Gr2 -).

2 Description des sujets de projet et de synthèse bibliographique

Vous trouverez ci-dessous la liste des synthèses bibliographiques proposés.

2.1 Synthèses bibliographiques

Reconnaissance d'expressions faciales (personne, genre, émotion, ...).

- B. Fasel, F. Monay, D. Gatica-Perez. *Latent semantic analysis of facial action codes for automatic facial expression recognition*. MIR04 Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval, page 181-188, 2004.
- Y. Tian T. Kanade, J. F. Cohn. *Comprehensive database for facial expression analysis*. In Proc. Int. Conf. Face and Gesture Recognition (FGR00), pages 46-53, 2000.

Modélisation computationnelle des émotions humaines.

- P. Gebhard, *ALMA - a layered model of affect*, in Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-05), pp. 29-36, 2005.
- Becker-Asano, C., Wachsmuth, I., *Affect simulation with primary and secondary emotions.*, In International Conference on Intelligent Virtual Agents, pp. 15-28, Springer, Berlin, Heidelberg, 2008.

Reconnaissance de gestes et/ou de postures.

- M.C. Roh, B. Christmas, J. Kittler, S.W. Lee, *Robust player gesture spotting and recognition in low-resolution sports video*, Computer Vision - ECCV, pp.347-358, 2006.
- R. Ibañez, Á. Soria, A. Teyseyre, M. Campo, *Easy gesture recognition for Kinect*, Advances in Engineering Software, volume 76, pp.171-180, 2014.

Détection d'émotion dans la voix.

- A. Mencattini, E. Martinelli, G. Costantini, M. Todisco, B. Basile, M. Bozzali et C. Di Natale., *Speech emotion recognition using amplitude modulation parameters and a combined feature selection procedure*. Knowledge-Based Systems, Volume 63, pp. 68-81, 2014.
- M. Tahon, L. Devillers, *Towards a small set of robust acoustic features for emotion recognition : challenges*, Transaction in Speech, Audio and Language Processing, volume 54, pp. 16-48, 2016.

Perception de scènes.

- I. Kajo, N. Kamel, Y., Ruichek, A. S. Malik, *SVD-Based Tensor-Completion Technique for Background Initialization*, IEEE Transactions on Image Processing, 27(6), pp. 3114-3126, 2018.
- Ilg, Eddy, et al., *Flownet 2.0 : Evolution of optical flow estimation with deep networks.*, IEEE conference on computer vision and pattern recognition (CVPR). Vol. 2. 2017.

Raisonnement spatio-temporel.

- Q. Cohen-Solal, M. Bouzid, A. Niveau, *Temporal Sequences of Qualitative Information : Reasoning about the Topology of Constant-Size Moving Regions*, IJCAI 2017 : 986-992
- L. Iocchi L., M. T. Lázaro, L. Jeanpierre, A. Mouaddib, *Personalized Short-Term Multi-modal Interaction for Social Robots Assisting Users in Shopping Malls*, ICSR 2015 : 264-274.

Navigation (en robotique) sociale.

- A. Mouaddib, *Vector-Value Markov Decision Process for multi-objective stochastic path planning*, International Journal of Hybrid Intelligent Systems, 9(1) : 45-60, 2012.
- K. H. Zilberstein, A. Mouaddib, *Multi-Objective MDPs with Conditional Lexicographic Reward Preferences*, AAAI 2015, pp. 3418-3424, 2015.

Transcription automatique (ou reconnaissance de la parole).

- F. Seide, G. Li, D. Yu. *Conversational speech transcription using context-dependent deep neural networks*. In Annual Conference of the Int. Speech Communication Association, pp. 152-158, 2011.
- P. C. Woodland, J. J. Odell, V. Valtchev, Steve J. Young, *Large Vocabulary Continuous Speech Recognition Using HTK*, In Proceedings of IEEE Int. Conf. on Acoustics, Speech and Signal Processing, pp. 125-128, 1994.

Opinion Mining et Analyse de sentiments.

- C. J. Hutto and E. Gilbert, *Vader : A parsimonious rule-based model for sentiment analysis of social media text.*, in Eighth International AAAI Conference on Weblogs and Social Media, 10p., 2014.
- T. Wilson, J. Wiebe, and P. Hoffmann, *Recognizing contextual polarity in phrase-level sentiment analysis*. In Proceedings of the conference on human language technology and empirical methods in natural language processing, pages 347-354. Association for Computational Linguistics, 2005.

Systèmes de dialogues.

- Frampton, M., Lemon, O. : *Recent research advances in reinforcement learning in spoken dialogue systems.*, The Knowledge Engineering Review 24, pp. 375-408, 2009.
- Allen, J., Ferguson, G., Miller, B., Ringger, E., Sikorski-Zollo, T. : *Dialogue systems : From theory to practice in Trains-96*. In : Handbook of Natural Language Processing, pp. 347-376, 2000.

Reconnaissance d'activité.

- S. Sigg, S. Haseloff, and K. David, *An Alignment Approach for Context Prediction Tasks in UbiComp Environments.*, IEEE Pervasive Computing, 9(4) : 90-97, 2010.

- D. Ashbrook and T. Starner, *User movement with GPS. Learning significant locations and predicting.*, in Proceedings of the 6th IEEE International Symposium on Wearable Computers, pages 101-108, Seattle, WA, USA, 2002.

Recherche d'information personnalisée

- Awadallah, A.H., Chu, W., He, X., Song, Y., White, R.W., Wang, H., *Enhancing personalized search by mining and modeling task behavior*, in Proceedings of WWW conference, pp. 1411-1420, 2013
- F. Liu, C. Yu, and W. Meng, *Personalized Web search for improving retrieval effectiveness.*, IEEE Transactions on knowledge and data engineering, volume 16, pp. 28-40, 2004.

Oculométrie et suivi de regards

- A Maria Feit, S Williams, A Toledo, A Paradiso, H Kulkarni, S Kane, M Ringel Morris, *Toward everyday gaze input : Accuracy and precision of eye tracking and implications for design*, Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 1118-1130, 2017.
- Dario D. Salvucci, Joseph H. Goldberg, *Identifying fixations and saccades in eye-tracking protocols*, in Proceedings of the 2000 symposium on Eye tracking research and applications, pp. 71-78, 2000.

Suivi d'objets dans une scène

- Andrea Cavallaro and Olivier Steiger and Touradj Ebrahimi, *Tracking video objects in cluttered background*, A Survey on Behaviour Analysis in Video Surveillance Applications, pp. 575-584, 2005.
- A. L. Gilbert, M. K. Giles, G. M. Flachs, R. B. Rogers and U. Y. Hsun, *A Real-Time Video Tracking System*, in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-2, no. 1, pp. 47-56, 1980.

Détection de communautés dans les réseaux sociaux

- S Fortunato, *Community detection in graphs*, Physics reports, Volume 486, Issue 3, pp. 75-174, 2010.
- Dinh, T. N. and Thai, M. T., *Community detection in scale-free networks : Approximation algorithms for maximizing modularity.*, Journal on Selected Areas in Communications, 31(6), pp. 997-1006, 2013.

Synthèse (visuelle et sonore) d'émotions

- M. Courgeon, C. Clavel, *MARC : a Framework that Features Emotion Models for Facial Animation during Human-Computer Interaction*, in Journal on Multimodal User Interfaces, Volume 7(4), pp. 311-319, 2013
- Marc Schröder, *Emotional Speech Synthesis - A Review*, in Proceedings of the 7th European Conference on Speech Communication and Technology (EUROSPEECH'01), pp. 561-564, 2001.

Environnements virtuels numériques pour l'apprentissage humain

- C Forsyth, A Graesser, A Olney, K Millis, B Walker and Z Cai, *Moody Agents : Affect and Discourse during Learning in a Serious Game*, in International Conference on Artificial Intelligence in Education (AIED2015), pp. 135-144, 2015.
- AS. Najar, A. Mitrovic, BM McLaren, *Learning with intelligent tutors and worked examples : selecting learning activities adaptively leads to better learning outcomes than a fixed curriculum*, User Modeling and User-Adapted Interaction Journal (UMUAI), volume 26, pp. 459-491, 2016.

N'importe quel sujet de votre choix, que je valide !

2.2 Projets

2.2.1 Modèle de cohésion de groupe en temps réel

Ce projet s'inscrit dans le cadre d'un ensemble d'expérimentations impliquant des interactions de groupes hybrides (groupes composés d'humains et d'agents virtuels). L'hypothèse fondamentale de ces travaux de recherche est que l'estimation de la cohésion du groupe au cours de l'interaction doit permettre d'améliorer celle-ci.

L'objectif est de développer un modèle permettant d'estimer la cohésion d'un groupe en temps réel à partir de jeux de données annotés pour la cohésion et pour d'autres signaux sociaux (transcription de la parole, prosodie, émotions, regards...).

A partir de jeux de données annotés fournis, le but de ce projet est de développer un modèle permettant d'estimer, à intervalles de 2 minutes, la cohésion du groupe et de prédire son évolution pour les deux minutes suivantes en prenant en compte l'historique des interactions du groupe. Il convient de prendre en compte les contraintes suivantes :

- de prédire la cohésion pour des situations variées : des groupes ayant un nombre de participants différents ou interagissant dans des contextes différents, ce qui implique de pouvoir prédire la cohésion même si toutes les données d'entrées théoriques ne sont pas disponibles (donc en utilisant un sous ensemble uniquement).
- de prédire la cohésion en ayant un quantité de données limitée.

Le projet inclut les tâches suivantes :

- Se documenter sur les jeux de données proposés et les annotations disponibles pour préparer les données d'entrée du modèle, c.-à-d. gérer les données manquantes, les données aberrantes, normaliser les données, encoder les variables catégorielles en numériques si besoin, regrouper des données pour créer de nouvelles données (ex : agréger dans un vecteur, moyenner des données individuelles pour représenter le groupe).
- Se documenter sur les modules/blocs/layers qui existent pour élaborer un modèle qui satisfasse l'ensemble des contraintes énoncées dans « description »
- Développer le modèle pour surpasser la baseline (reshma) puis démontrer la robustesse du modèle aux situations variées.

2.2.2 Pilotage d'un avatar par mise en miroir

Ce projet s'inscrit dans le cadre d'un ensemble d'expérimentations avec des compagnons artificiels, agents virtuels ou robotiques.

L'objectif est de développer et implémenter un modèle capable de reproduire en temps réel les comportements d'un utilisateur humain pour animer un agent virtuel. A partir de deux dispositifs de captation, à savoir, une caméra et un microphone, l'objectif de ce projet est de tirer parti des récentes avancées afin de détecter : 1) la posture (position et orientation de la tête, position des mains et du corps dans l'espace ; 2) les expressions faciales et 3) la voix (prosodie et transcription de la parole) d'un être humain pour ensuite reproduire fidèlement l'ensemble de ses comportements en temps réel sur un avatar.

Le projet inclut les tâches suivantes :

- Réaliser un inventaire des méthodes/programmes/outils qui peuvent être utilisés pour détecter la posture, les expressions faciales et la voix (prosodie et transcription)
- Permettre l'utilisation du/des programmes en temps réel en exploitant un flux vidéo et audio
- Transposer les comportements humains détectés sur un avatar dans la plateforme Unity3D.

2.2.3 Détection de regard mutuel et Focus d'attention

Ce projet s'inscrit également dans le cadre d'un ensemble d'expérimentations avec des compagnons artificiels, agents virtuels ou robotiques.

Dans le cadre d'une communauté mixte, composée à la fois d'agents artificiels (virtuels et/ou robots) et d'humain, l'objectif est, au cours d'une interaction d'identifier avec précision les regards mutuels :

- pour un robot/agent avec un interlocuteur humain,
- entre deux autres interlocuteurs, qu'ils soient humains ou robot/agent ;

ainsi que les focus d'attention, c'est-à-dire lorsque deux interlocuteurs regardent le même objet/endroit.

Le projet inclut les tâches suivantes :

- Réaliser un inventaire des méthodes/programmes/outils qui peuvent être utilisés pour faire du suivi de regard à partir d'une webcam ;
- Permettre l'utilisation du/des programmes en temps réel en exploitant un flux vidéo
- Exploiter ces outils pour détecter les regards mutuels et focus d'attention.