

APPRENTISSAGE DE REPRÉSENTATIONS MULTIMODALES POUR LA SÉMANTISATION DES PARAMÈTRES DU MOUVEMENT

Établissement : IMT Mines Alès (École nationale supérieure des mines d'Alès)

Centre et équipe de recherche : Équipe I3A (Informatique, Image et Intelligence Artificielle)
LGI2P (Laboratoire de Génie Informatique et d'Ingénierie de Production)

Localisation : Alès (1h de Montpellier, 45 min de Nîmes)

École doctorale : I2S Information Structures Systèmes – <http://www.edi2s.univ-montp2.fr>

Spécialité : Informatique (section CNU 27) / Computer Science

Encadrement : Sylvie Ranwez, Professeur à IMT Mines Alès ; Andon Tchechmedjiev, Maître de Conférences à l'IMT Mines Alès ; Julien Lagarde Maître de Conférences à l'Université de Montpellier

Financement : Allocation doctorale de la région Occitanie avec co-financement IMT Mines Alès

Début de la thèse : Automne 2020

Date limite de candidature : 6 juillet 2020

Contact : sylvie.ranwez@mines-ales.fr

andon.tchechmedjiev@mines-ales.fr

Résumé

Le mouvement humain est régi par un ensemble complexe de mécanismes, et fait partie intégrante de la vie de chacun, que ce soit pour communiquer, travailler, produire de l'art ou encore pour s'adonner à des activités physiques. Comprendre ces mécanismes ainsi que l'influence du contexte (physique, sociétal, culturel) sur la signification ou la description du mouvement, est essentielle pour nombre d'applications liées à l'analyse du mouvement : de la conception de représentations artistiques, à la rééducation fonctionnelle, ou encore l'optimisation de la performance sportive, par exemple. Dans ce projet de thèse, nous souhaitons étudier comment contextualiser et donner du sens aux paramètres du mouvement, qu'ils soient directs (coordonnées de points de référence) ou non (signaux neurophysiologiques) en les croisant avec des modèles de connaissances (taxonomies, ontologies) de domaine et/ou des modèles de langage contextualisés (plongements de graphes et modèles Transformers pré-entraînés) grâce à la mise en œuvre de techniques d'apprentissage de représentations multimodales profondes. Nous envisageons des applications à l'optimisation de la performance sportive et au diagnostic de pathologies du mouvement (maladie de Parkinson, lombalgie).

Mots-clés

Représentations paramétriques du mouvement, ingénierie des connaissances (taxonomies, ontologies), traitement automatique du langage naturel, apprentissage automatique de représentations multimodales, reconnaissance d'activité.

Description du sujet de thèse

L'objectif du projet SemTaxM (Sémantique et Taxonomie du Mouvement) est d'étudier les taxonomies liées à la classification du mouvement (au sens de la classification des objets d'études) afin de faire émerger un modèle général du mouvement au travers de l'identification de phénomènes de composition (émergence d'une action complexe à partir d'actions simples) et d'invariance dans les paramètres du mouvement (invariance temporelle, d'échelle). Ce sujet de thèse, « Apprentissage de représentations multimodales pour *sémantiser* les paramètres du mouvement », constitue un premier pas vers cet objectif, en explorant les techniques d'apprentissage de représentations multimodales pour la *sémantisation* des paramètres du mouvement (paramètres physiques directs issus d'estimation de pose ou de capteurs, ou paramètres neurophysiologiques) grâce au croisement avec des représentations de connaissances (plongements de taxonomies ou de graphes de connaissance, modèles de langue contextualisés pré-entraînés).

La première dimension explorée sera celle de l'association des représentations paramétriques directes du mouvement avec soit une taxonomie de concepts, soit une description du mouvement en langage naturel par des expressions courtes. Dans l'état de l'art, (Ahuja et Morency, 2019) ont proposé un système qui permet d'apprendre une représentation par plongement combinant un corpus de poses (poses successives d'un squelette dans le temps) et un modèle de plongements lexicaux classique (Word2Vec, Mokolov et al. 2012). Cette approche pourra servir de base de comparaison pour nos modèles, qui intégreront des représentations vectorielles pré-entraînés (Bert/Camembert/Flaubert ; Devlin et al. 2019) ou des plongement de graphes de connaissance. Nous pourrions exploiter de nombreux jeux de données publics de classification d'activités, notamment de vidéos accompagnées d'instructions (Shruti et al. 2019, Shi et al. 2019, Wang et al. 2019) que nous associerons à des systèmes d'estimation de pose 3D (par exemple VideoPose3D de Facebook Research ; Pavllo et al. 2019). Nous n'imposerons pas de contraintes sur le paradigme d'entraînement utilisé pour les modèles, étant donné des résultats encourageants des approches d'apprentissage par renforcement dans l'état de l'art, pour l'imitation de mouvement (Peng et al. 2018) ou l'apprentissage de politiques conditionnées par le langage (Zhong et al. 2020).

La deuxième dimension explorée par le sujet sera l'utilisation de paramètres indirects du mouvement tel que la fréquence cardiaque, la conductance cutanée, l'électromyographie (EMG), électrocardiographie (ECG), électroencéphalographie (EEG), NIRS) pour construire des modèles d'activité sémantisés (Ramoly et al. 2017). Nous pourrions ainsi étudier l'extension des modèles développés lors de la première étape. Les cas d'application envisagés sont l'optimisation de la performance sportive et la rééducation fonctionnelle. La disponibilité de données terrain ne peut être garantie à ce stade, mais il sera possible de faire usage de la plateforme de capture du mouvement IMT Mines Alès/EuroMov DHM pour constituer des jeux de données de travail.

Objectifs et résultats attendus

Dans un premier temps la production d'un modèle de représentations profondes multimodales permettant de sémantiser des représentations paramétriques du mouvement à l'aide d'ontologies ou de modèles de langue contextualisés. Dans un deuxième temps nous envisagerons l'extension de cette approche à des mesures indirectes du mouvement (mesures neurophysiologiques). L'évaluation se fera sur deux cas d'application : diagnostic de parkinsoniens à partir de vidéos et reconnaissance d'activités pour l'optimisation de la performance sportive.

Le financement de thèse s'accompagne d'un budget valorisation, qui permettra la publication de résultats méthodologiques dans des conférences internationales de rang A/A+ (par ex. *ACL, ISWC, ICLR). Nous envisageons une publication journal portant sur le cas d'application lien au diagnostic des pathologies du mouvement (par ex. JBI ou JBS).

Ressources de calcul et infrastructure

Le LGI2P possède plusieurs machines de calcul avec GPU qui pourront être mises à disposition du candidat sélectionné pour les expérimentations. Cela se verra complété par un accès au cluster Jean Zay du CNRS à hauteur de 10000h de calcul GPU sur 4 cartes NVIDIA Quadro V100.

Précisions sur l'encadrement

Directrice de thèse : Sylvie Ranwez, Professeur, LGI2P/IMT Mines Alès,
sylvie.ranwez@mines-ales.fr, , 04 34 24 62 62

Encadrant de proximité : Andon Tchechmedjiev, MCF, LGI2P/IMT Mines Alès
andon.tchechmedjiev@mines-ales.fr, 04 34 24 62 16

Encadrant expertise mouvement humain : Julien Lagarde, MCF HDR, Université de Montpellier
julien.lagarde@umontpellier.fr

Expertise domaine santé : Valérie Cochen de Cock, Neurologue, Clinique Beau Soleil
valerie.cochen@gmail.com

Sylvie Ranwez est une experte en représentation des connaissances et utilise les ontologies en particulier pour la recherche d'informations, l'indexation et la visualisation de données via des cartes sémantiques. **Andon Tchechmedjiev** se spécialise en sémantique computationnelle, traitement automatique du langage naturel et en représentation de connaissances, avec un intérêt particulier pour les applications en informatique médicale et en sciences du mouvement humain. Il anime l'axe transverse Sémantique et Taxonomie du Mouvement de l'équipe I3A et de la future UMR Euromov DHM. **Julien Lagarde** se spécialise autour de la problématique de synchronisation dans le mouvement humain au sein du laboratoire Euromov de l'université de Montpellier et s'intéresse de près aux représentations paramétriques du mouvement et à la taxonomie du mouvement. **Valérie Cochen de Cock** est neurologue à la clinique Beau Soleil à

Montpellier et chercheuse au laboratoire Euromov. Elle se spécialise dans les troubles du sommeil et du mouvement humain, avec un intérêt particulier pour la maladie de Parkinson.

Profil du candidat

Master II en Informatique, ingénieur informaticien ou équivalent

Une bonne maîtrise des techniques utilisées dans les domaines de l'apprentissage automatique et du traitement automatique du langage naturel ou du traitement d'images est souhaitée, en particulier par l'utilisation d'approche par apprentissage profond. Des connaissances en ingénierie des connaissances (ontologies) / web sémantique sont attendues.

Des compétences en développement informatique et en particulier dans le développement de codes de calcul (algorithmes) sont requises (python, pytorch ou tensorflow). Le candidat aura un bon niveau en anglais (notamment à l'écrit), ainsi qu'une bonne capacité à travailler en équipe et à respecter les échéances.

Présentation de l'établissement et du centre d'accueil

L'**institut Mines-Télécom (IMT)** est un grand établissement public dédié à l'enseignement supérieur et la recherche pour l'innovation ; c'est le premier groupe d'écoles d'ingénieurs en France. Acteur majeur du croisement entre les sciences et les technologies du numérique et de l'ingénieur, l'IMT met les compétences de ses écoles en perspective dans les grands champs des transformations numériques, industrielles, énergétiques et éducatives. L'IMT fédère 11 écoles d'ingénieur publiques réparties sur le territoire national, qui forment 13 500 ingénieurs et docteurs. L'IMT emploie 4500 femmes et hommes et dispose d'un budget annuel de la recherche partenariale de 400M€ au sein de 55 centres de recherche rattachés à ses écoles. Il produit chaque année plus de 2000 publications et 60 brevets.

IMT Mines Alès est une des écoles de l'IMT. Forte de 180 ans d'histoire au service de la science et des entreprises, l'école dispose d'une solide culture scientifique et technique qu'elle met au service de l'enseignement, de la recherche et du transfert technologique. L'école emploie 350 personnes et forme plus de 1100 élèves ingénieurs et chercheurs ouverts sur le monde. Ses 3 centres de recherche développent une activité de haut niveau scientifique en partenariat avec les entreprises, dans les domaines de l'environnement, des risques, des matériaux, du génie civil, de l'intelligence artificielle et du génie industriel et numérique. Les valeurs promues à l'école sont l'audace, l'engagement, le partage et l'excellence.

L'équipe I3A (informatique, image et intelligence artificielle) est une des deux équipes du centre de recherche et d'enseignement LGI2P. Forte de son expertise scientifique dans différents champs disciplinaires de l'intelligence artificielle et de l'ingénierie logicielle, elle contribue à la définition de méthodes et de solutions logicielles pour accompagner l'Homme dans sa relation à un système complexe. Héritière de ses travaux sur l'automatisation cognitive, sa recherche actuelle est menée dans le cadre d'une unité mixte DHM (digital health in motion) avec le Centre EuroMov de l'Université de Montpellier. Cette nouvelle unité de recherche, ambitionne une

ligne de recherche translationnelle sur « La Santé Numérique en Mouvement ». Sa recherche porte sur l'étiologie du mouvement humain, vu comme le niveau d'intégration des phénomènes biologiques et cognitifs au cours de nos échanges informationnels permanents avec l'environnement. Les sciences du mouvement et de la santé reposent très largement sur l'analyse de données hétérogènes et sur la prise en compte de connaissances métier (médicales notamment). Le positionnement de l'équipe I3A propose des approches hybrides basées sur quatre piliers scientifiques :

- Intelligence artificielle : représentation des connaissances et apprentissage automatique,
- Aide à la décision et optimisation,
- Imagerie numérique,
- Ingénierie système et logicielle dirigée par les modèles.

I3A est organisée en 3 thèmes et 2 axes transversaux :

- MIB – Monitoring and Improving Behaviour
- PIAS – Perception In Action and Synchronisation
- LAC – Learning and Complexity
- SemTaxM – Sémantique et Taxonomie du Mouvement [Axe de recherche transverse]
- Factory [Axe transverse]

Cette thèse correspond aux travaux sur l'axe transversal Sémantique et Taxonomie du Mouvement (SemTaxM), mais est administrativement positionnée sur le thème Learning And Complexity (LAC).

Références bibliographiques

- C. Ahuja and L. Morency, "Language2Pose: Natural Language Grounded Pose Forecasting," *2019 International Conference on 3D Vision (3DV)*, Québec City, QC, Canada, 2019, pp. 719-728. doi: 10.1109/3DV.2019.00084
- I. Chami, S. Abu-El-Haija, B. Perozzi, C. Ré, K. Murphy. **Machine Learning on Graphs: A Model and Comprehensive Taxonomy**. 2020.1–38. <http://arxiv.org/abs/2005.03675>.
- J. Devlin, M.W. Chang, K. Lee, K. Toutanova. **BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding**. 2019. 1–16. <https://arxiv.org/abs/1810.04805v2>
- S. Palaskar, J. Libovický, S. Gella, and F. Metze. "Multimodal Abstractive Summarization for How2 Videos." arXiv preprint arXiv:1906.07901 (2019).
- D. Pavlo, C. Feichtenhofer, D. Grangier, and M. Auli. "3D human pose estimation in video with temporal convolutions and semi-supervised training." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7753-7762. 2019.
- X. Bin Peng, A. Kanazawa, J. Malik, P. Abbeel, and S. Levine. "SFV: Reinforcement learning of physical skills from videos." *ACM Transactions on Graphics (TOG)* 37, no. 6 (2019): 178.

- N. Ramoly, V. Vassout, |A. Bouzeghoub, M. El Yacoubi, M. Hariz. **Refining visual activity recognition with semantic reasoning**. IEEE 31st International Conference on Advanced Information Networking and Applications (AINA 2017), Mar 2017, Taipei, Taiwan. pp.720 - 727, ff10.1109/AINA.2017.98ff. fhal-01587280f
- B. Shi, L. Ji, Y. Liang, N. Duan, P. Chen, Z. Niu, and M. Zhou. "**Dense procedure captioning in narrated instructional videos**." In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 6382-6391. 2019.
- W. Wang, Y. Wang, S. Chen, and Q. Jin. "**YouMakeup: A Large-Scale Domain-Specific Multimodal Dataset for Fine-Grained Semantic Comprehension**." In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 5136-5146. 2019.
- X. Yang, P. Ramesh, R. Chitta, S. Madhvanath, E.A. Bernal, and J. Luo. "**Deep multimodal representation learning from temporal data**." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5447-5455. 2017.
- V. Zhong, T. Rocktäschel, E. Grefenstette. **RTFM: Generalising to New Environment Dynamics via Reading**. International Conference on Learning Representations. 2020.

Processus de candidature

Le dossier de candidature doit être envoyé par mail aux encadrants principaux : sylvie.ranwez@mines-ales.fr et andon.tchechmedjiev@mines-ales.fr avant le 6 juillet 2020 (5 Juillet 2020, 23:59:59, n'importe-où sur terre). Il sera composé de :

- Un CV détaillant en particulier les expériences dans les domaines en lien avec le sujet de thèse.
- Un document d'intention et de motivation identifiant un placement par rapport à l'état de l'art et présentant la vision sur le sujet du candidat (3 pages max).
- Les relevés de notes et les classements obtenus dans les dernières années de formation (master ou équivalent).
- Des lettres de recommandation.

Les candidats dont le dossier sera retenu seront invités à passer un entretien où ils feront une présentation de 15 minutes, suivie de 15 minutes de questions. Nous encourageons les candidats potentiels à nous contacter en amont pour avoir des informations complémentaires pouvant les aider dans la préparation de leurs dossiers.