Offre de thèse en informatique
(Machine Learning, LLM, Graphe)

Titre : Combinaison LLM et GNN pour la fusion de représentations multimodales : Application à l'extraction d'information dans les données semi-structurées liées aux activités humaines ou professionnelles.

Mots-clés
Graph, Neural Networks, LLM, fusion multi-modale, données semi-structurées, RAG, frugal AI, DocVQA, extraction d'information.

Encadrants
David TELISSON, Jean-Yves RAMEL
Laboratoire LISTIC 7 Université Savoie-Mont-Blanc

Contexte scientifique

Les données générées dans les contextes professionnels ou issus des activités humaines sont de plus en plus riches, volumineuses et hétérogènes. Elles se présentent souvent sous une forme semi-structurée : textes libres, images, documents administratifs, formulaires numérisés, tableaux, métadonnées, etc. Cette diversité rend les tâches d'extraction et de structuration de l'information particulièrement complexes, en particulier sous des contraintes de temps ou de ressources matérielles. Les approches récentes en intelligence artificielle, notamment les modèles de langage de grande taille (LLMs), les Graph Neural Networks (GNNs), et les méthodes de fusion de données offrent de nouvelles perspectives pour relever ces défis.

Objectifs de la thèse

L'objectif principal est de concevoir des modèles hybrides frugaux capables de simplifier l'extraction d'informations structurées à partir de données multi-modales et semi-structurées issues de contextes métiers ou humains. La thèse s'inscrit dans le cadre du projet FONDUE, qui vise à construire des méta-modèles FAIR (Findable, Accessible, Interoperable, Reusable) et exécutables sur du matériel standard d'extraction d'information et d'aide à la décision dans les données semi-structurées liées aux activités humaines ou professionnelles.

Plus précisément, le doctorant ou la doctorante devra :

- Explorer les méthodes de fusion multi-modale permettant de combiner texte, image, tableau et métadonnées dans une représentation unifiée.
- Représenter les informations semi-structurées sous forme de graphes et exploiter les GNNs pour le raisonnement ou la classification.
- Développer des méthodes pour enrichir les données extraites et stockées dans les graphes sémantiques.

- Proposer des architectures légères combinant modèles de langage (LLMs) et GNNs pour limiter les coûts de calcul.
- Évaluer ces approches sur des cas d'usage concrets comme le Document Visual Question Answering (DocVQA) ou la reconnaissance d'information dans des documents administratifs scannés.

Méthodologie proposée

- Etude de l'état de l'art (cf références, https://github.com/aolivtous/LLMs-for-DocVQA)
- Mise en œuvre et expérimentation de "pipelines monomodaux" de traitement des données semi-structurées (OCR + LLM, Analyse de layout/mages + GNN).
- Développement de modèles combinant LLMs et GNNs pour la représentation et la classification d'information semi-structurée multimodales.
- Expérimentation de stratégies de fusion sans re-entraînement complet, notamment via des techniques de *model ensembling* ou d'injection de connaissances.
- Évaluation empirique des performances (précision, coût, robustesse) sur des jeux de données publics et des cas industriels (le cas échéant).

Profil recherché

- Master en intelligence artificielle, informatique, data science ou domaine connexe.
- Compétences solides en machine learning, NLP ou traitement d'images.
- Intérêt pour les approches multi-modales et les architectures hybrides.
- Maîtrise de Python, Pytorch/Tensorflow, et des bibliothèques LLM (Transformers, LangChain, etc.).

Encadrement et contexte

La thèse sera encadrée par David Télisson et JY Ramel au sein de l'équipe ReGards du LISTIC, dans le cadre de la chaire MIAI FONDUE (https://miai-cluster.univ-grenoble-alpes.fr/) L'environnement de recherche inclut un accès à des plateformes de calcul, des données réelles et un écosystème académique et industriel dynamique.

Envoyer CV + lettre de motivation + recommandation à :
jean-yves.ramel[at]univ-smb.fr et/ou david.telisson[at]univ-smb.fr

Références bibliographiques

Wang, Dongsheng and Ma, Zhiqiang and Nourbakhsh, Armineh and Gu, Kang and Shah, Sameena, DocGraphLM: Documental Graph Language Model for Information Extraction, Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval SIGIR '23 url = https://arxiv.org/html/2401.02823v1 {https://doi.org/10.1145/3539618.3591975}

Yu Wang, Nedim Lipka, Ryan A. Rossi, Alexa Siu, Ruiyi Zhang, Tyler Derr, Knowledge Graph Prompting for Multi-Document Question Answering, arXiv:2308.11730v3 [cs.CL] 25 Dec 2023

Guo, Taicheng and Chen, Xiuying and Wang, Yaqi and Chang, Ruidi and Pei, Shichao and Chawla, Nitesh V. and Wiest, Olaf and Zhang, Xiangliang, Large Language Model Based Multi-agents: A Survey of Progress and Challenges, IJCAI24 (2024) https://doi.org/10.24963/ijcai.2024/890},

Lewis, Patrick and Perez, Ethan and Piktus, Aleksandra and Petroni, et Al, Retrieval-Augmented Generation for Knowledge-Intensive NLP Task, NEURIPS 2020 volume 33, p9459--9474, (2020).

Deyao Zhu, Jun Chen, Xiaoqian Shen, Xiang Li, Mohamed Elhoseiny, MiniGPT-4: Enhancing Vision-Language Understanding with Advanced Large Language Models, ICLR 2024

M. Mathew, D. Karatzas and C. V. Jawahar, DocVQA: A Dataset for VQA on Document Images, *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 2199-2208 (2021), doi: 10.1109/WACV48630.2021.00225

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., ... & Sun, M. (2020). Graph neural networks: A review of methods and applications. *AI open*, *1*, 57-81. IEEE TNNLS 2020.

Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. Yu. A Comprehensive Survey on Graph Neural Networks *IEEE Transactions on Neural Networks and Learning Systems, 32 (1): 4-24* (*2021*) doi:10.1109/TNNLS.2020.2978386

Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18