

Spatial and Temporal Reasoning with LLMs for Natural Language Comprehension and Grounding

Parisa Kordjamshidi

Michigan State University, USA

Recent research in Natural Language Processing (NLP) has revealed that deep learning models, particularly large language models (LLMs) trained on huge amounts of data suffer from a lack of interpretability and generalizability. This issue extends to spatial and temporal reasoning over natural language and visual data too. Although LLMs can impress us by fluently generating articles given a prompt, they often fail in basic reasoning tasks like understanding that "left" is the opposite of "right." Real-world problem-solving requires computational models that involve multiple interdependent learners, extensive composition, and reasoning based on additional knowledge beyond the available data. Our research endeavors at the Heterogeneous Learning and Reasoning Lab (HLR)¹ focus on tackling some of these challenges.

In the first part of my talk, I will discuss our recent research on three key areas. Firstly, we have evaluated the spatial reasoning capabilities of large language models over text and introduced new benchmarks specifically designed for this purpose [1, 2]. Secondly, we have developed architectures capable of capturing spatial and temporal information about entities and their activities, enabling procedural reasoning [3, 4]. Lastly, for vision and language grounding and navigation, we have developed new modules integrated with large vision and language model backbones. We pre-train these modules with novel synthesized indirect supervision resources to capture fine-grained semantics required for an accurate and explainable instruction following and navigation in a visual environment [5, 6, 7].

In the second part of my talk, I will introduce DomiKnowS, a Declarative learning-based programming framework. DomiKnowS is designed to facilitate the integration of learning and reasoning, leveraging both symbolic and sub-symbolic representations to solve complex AI-complete problems. This framework seamlessly integrates domain knowledge, represented symbolically as logical constraints, into deep models using various underlying algorithms. We cover a variety of training and inference time algorithms. Additionally, I will present GIUECons [8, 9], a new benchmark comprising tasks and models specifically designed for evaluating algorithms that aim to integrate logical constraints into deep models.

STRL '23: Second International Workshop on Spatio-Temporal Reasoning and Learning, 21 August 2023, Macao, S.A.R

✉ kordjams@msu.edu (P. Kordjamshidi)

🌐 <https://www.cse.msu.edu/~kordjams> (P. Kordjamshidi)

🆔 0000-0002-0877-7063 (P. Kordjamshidi)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

📄 CEUR Workshop Proceedings (CEUR-WS.org)

¹<https://hlr.github.io/>

References

- [1] R. Mirzaee, H. R. Faghihi, Q. Ning, P. Kordjamshidi, SpartQA: : A textual question answering benchmark for spatial reasoning, in: 2021 Annual Conference of the North American Chapter of the Association for Computational Linguistics, 2021. [arXiv:2104.05832](https://arxiv.org/abs/2104.05832).
- [2] R. Mirzaee, P. Kordjamshidi, Transfer learning with synthetic corpora for spatial role labeling and reasoning, in: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 2022, pp. 6148–6165. URL: <https://aclanthology.org/2022.emnlp-main.413>.
- [3] H. Rajaby Faghihi, P. Kordjamshidi, C. M. Teng, J. Allen, The role of semantic parsing in understanding procedural text, in: Findings of the Association for Computational Linguistics: EACL 2023, Association for Computational Linguistics, Dubrovnik, Croatia, 2023, pp. 1837–1849. URL: <https://aclanthology.org/2023.findings-eacl.137>.
- [4] H. Rajaby Faghihi, P. Kordjamshidi, Time-stamped language model: Teaching language models to understand the flow of events, in: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, Online, 2021, pp. 4560–4570. URL: <https://aclanthology.org/2021.naacl-main.362>. doi:10.18653/v1/2021.naacl-main.362.
- [5] Y. Zhang, P. Kordjamshidi, VLN-trans: Translator for the vision and language navigation agent, in: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Toronto, Canada, 2023, pp. 13219–13233. URL: <https://aclanthology.org/2023.acl-long.737>.
- [6] Y. Zhang, P. Kordjamshidi, LOViS: Learning orientation and visual signals for vision and language navigation, in: Proceedings of the 29th International Conference on Computational Linguistics (CoLING), International Committee on Computational Linguistics, Gyeongju, Republic of Korea, 2022, pp. 5745–5754. URL: <https://aclanthology.org/2022.coling-1.505>.
- [7] Y. Zhang, P. Kordjamshidi, Vision and language navigation agent with explanation ability, Under Review, 2023.
- [8] H. R. Faghihi, A. Nafar, C. Zheng, R. Mirzaee, Y. Zhang, A. Uszok, A. Wan, T. Premisri, D. Roth, P. Kordjamshidi, GLUECons: A generic benchmark for learning under constraints, in: Proceedings of Thirty-Seventh AAAI conference on artificial intelligence, accepted, 2023.
- [9] H. Rajaby Faghihi, Q. Guo, A. Uszok, A. Nafar, P. Kordjamshidi, DomiKnowS: A library for integration of symbolic domain knowledge in deep learning, in: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Association for Computational Linguistics, 2021, pp. 231–241.