How do Large Language Models Understand Trajectory Data? Insights from Various Trajectory Formats and Response Strategies for Transportation Mode Detection

Yingpeng LI

Introduction

The effectiveness of large language models (LLMs) in transportation mode detection remains underexplored, creating a significant research gap in understanding how these models process trajectory data: (1) Which trajectory formats are most effectively understood by LLMs? (2) How do different strategies impact TMD? (3) Do Chain-of-Thought (CoT)-guided responses lead to hallucinations? If so, what types of hallucinations are produced? This study used the Geolife dataset to investigate the ability of pre-trained (PT) and fine-tuned (FT) LLMs to detect transportation modes across 14 trajectory formats. Meanwhile, two response strategies are compared.

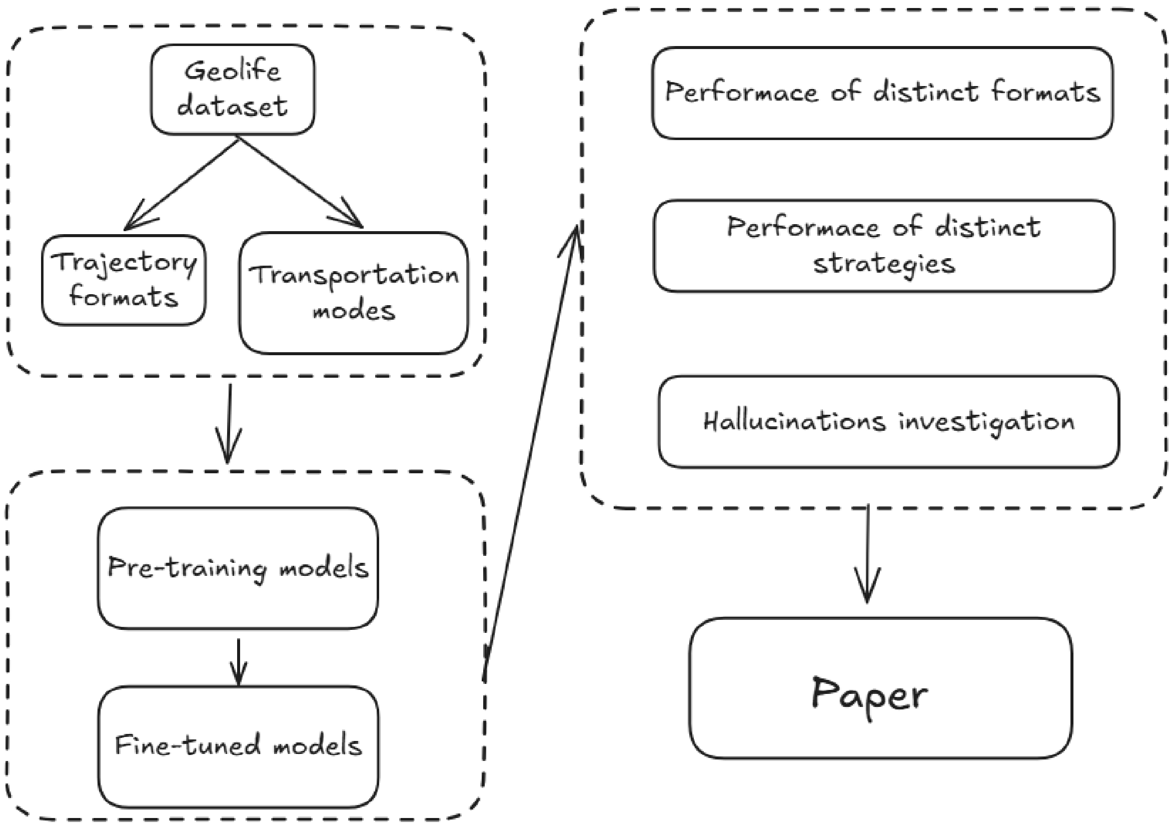

Figure 1. Mind map.

Methods

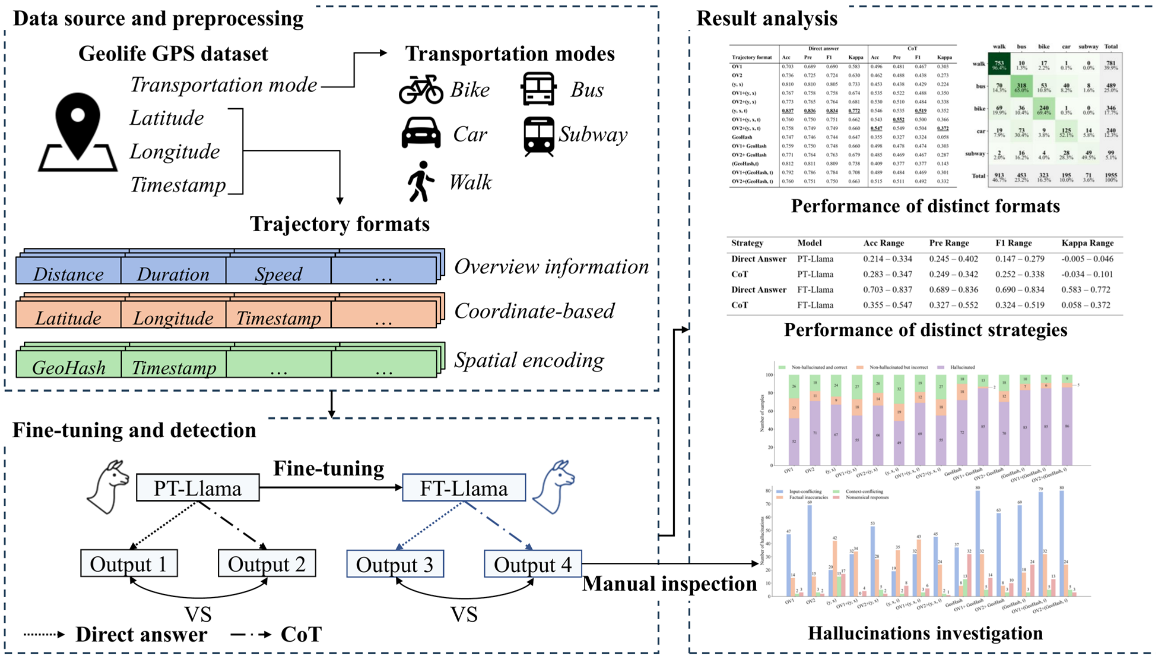

Figure 2. Flowchart.

Metrics

$$\text{Accuracy} = \frac{\sum_{i=1}^{N} TP_i}{\sum_{i=1}^{N}(TP_i + FP_i + FN_i)}$$

$$\text{Weighted F1-score} = \sum_{i=1}^{N} w_i \cdot \frac{2TP_i}{2TP_i + FP_i + FN_i}$$

$$\text{Cohen's Kappa} = \frac{P(\text{observed}) - P(\text{expected})}{1 - P(\text{expected})}$$

Results

Table 1: Performance evaluation of the FT-Llama in distinct formats.

Trajectory format	DA Acc	DA F1	DA Kappa	CoT Acc	CoT F1	CoT Kappa
Overview information 1	0.703	0.690	0.583	0.496	0.467	0.303
Overview information 2	0.736	0.724	0.630	0.462	0.438	0.273
(Lat, Lon)	0.822	0.818	0.751	0.460	0.429	0.238
OV1 + (Lat, Lon)	0.748	0.737	0.646	0.538	0.501	0.361
OV2 + (Lat, Lon)	0.758	0.748	0.660	0.524	0.489	0.340
(Lat, Lon, time)	0.852	0.849	0.793	0.447	0.415	0.212
OV1 + (Lat, Lon, time)	0.750	0.740	0.651	0.525	0.482	0.348
OV2 + (Lat, Lon, time)	0.790	0.783	0.706	0.508	0.482	0.325
GeoHash	0.747	0.744	0.647	0.355	0.324	0.058
OV1 + GeoHash	0.759	0.748	0.660	0.498	0.474	0.303
OV2 + GeoHash	0.771	0.763	0.679	0.485	0.467	0.287
(GeoHash, time)	0.812	0.809	0.738	0.409	0.377	0.143
OV1 + (GeoHash, time)	0.792	0.784	0.708	0.489	0.469	0.301
OV2 + (GeoHash, time)	0.760	0.750	0.663	0.515	0.492	0.332

Table 2: Performance comparison of PT-Llama and FT-Llama.

Strategy	Model	Acc Range	F1 Range	Kappa Range
Direct Answer	PT-Llama	0.214 – 0.334	0.147 – 0.279	−0.010 – 0.041
CoT	PT-Llama	0.295 – 0.347	0.264 – 0.338	−0.013 – 0.101
Direct Answer	FT-Llama	0.703 – 0.852	0.690 – 0.849	0.583 – 0.793
CoT	FT-Llama	0.355 – 0.538	0.324 – 0.501	0.058 – 0.361

Table 3: Examples of the four types of hallucinations.

Type	Response	Fact
Input-conflicting	{Reasoning}… the duration was only 116 seconds…	The duration was 616 seconds.
Factual inaccuracies	{Reasoning}… It reached the Tiananmen Square…	The trajectory did NOT reach the Tiananmen Square.
Context-conflicting	{Reasoning}… the mode is bus… but the final answer is car.	The final answer should be bus based on the reasoning process.

Type	Response	Fact
Nonsensical responses	{Nonsensical sentences}... The final answer is: walk.	The model did NOT provide contextually relevant reasoning.

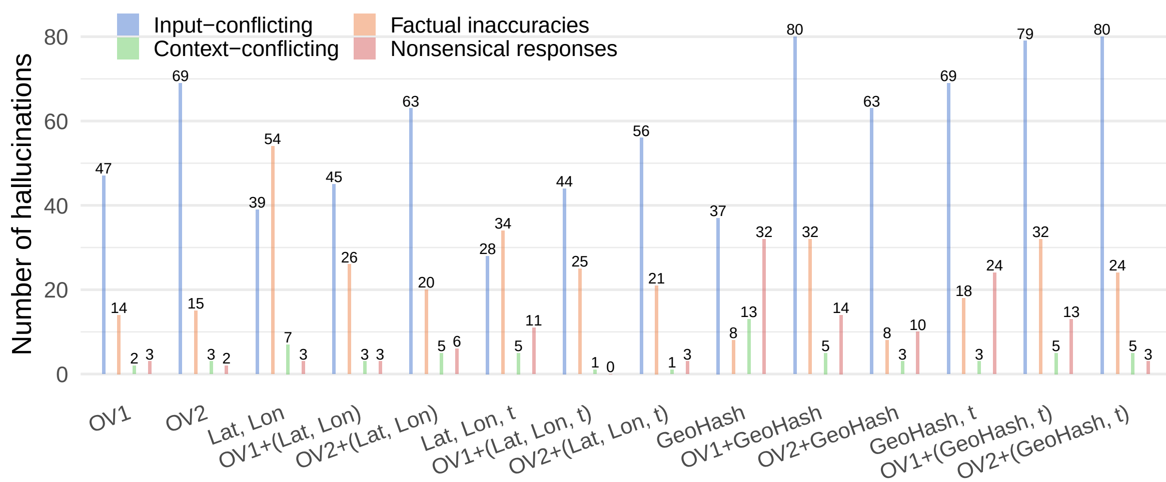

Figure 3. Number of hallucinated (100 samples/format, manual inspection).

Discussion

1. Effectiveness of the different trajectory formats.
2. Direct answer or CoT strategy: performance trade-offs.
3. Potential risks of hallucinations in fine-tuned models.
4. Limitations and future directions.

Conclusion

1. Classification performance varies according to trajectory format. The (Lat, Lon, time) format achieves better performance in direct-answer-guided FT-Llama (accuracy = 85.2%).
2. FT-Llama significantly outperforms PT-Llama, with the direct answer strategy yielding better classification results than the CoT strategy.
3. The CoT strategy introduces hallucinations.
4. In data-rich formats, there are more input-conflicting hallucinations and factual inaccuracies, while context-conflicting hallucinations and nonsensical responses are less frequent.

References

[1] Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., & Wang, Y. (2024). A survey on evaluation of large language models. ACM Transactions on Intelligent Systems and Technology, 15(3), 1-45. https://doi.org/10.1145/3641289

[2]

[...]

[51] Zheng, Y., Xie, X., & Ma, W.-Y. (2009, April). Mining Interesting Locations and Travel Sequences From GPS Trajectories Proceedings of International conference on World Wide Web 2009, https://www.microsoft.com/en-us/research/publication/mining-interesting-locations-and-travel-sequences-from-gps-trajectories/