Map Matching of Location Data Trajectories: A Heterogeneous and Bayesian-Optimized Hidden Markov Approach

A. Ruohan Li[a], B. Chenfeng Xiong[a*], C. Arash Tavakoli[a], D. C. Nataraj[a]

[a] Villanova University, 800 E Lancaster Ave, Villanova PA, 19085 U.S.
* Corresponding author: chenfeng.xiong@villanova.edu

Extended abstract submitted for presentation at the Conference in Emerging Technologies in Transportation Systems (TRC-30)
September 02-04, 2024, Crete, Greece

April 29, 2024

Keywords: Map Matching, Hidden Markov, Bayesian, Optimization, Location-Based Service Data

1 INTRODUCTION

Map matching is a pivotal process in transportation research, serving to align raw geographic location data with digital road network maps to accurately deduce the routes traversed by individuals or vehicles. This alignment is essential for reconstructing actual travel paths, providing critical insights for analyzing travel behavior, identifying preferred routes, and enhancing transportation system efficiency. However, map matching presents considerable challenges due to the complexity and variability of realistic road networks, such as the intricate layout of urban areas, overlapping, linked, and parallel roads on two-dimensional maps, and the inherent inaccuracies and frequency variations in geographic location data, particularly those from Location-Based Services (LBS) generated by mobile and other GPS-enabled devices.

The introduction of HMM (Newson and Krumm, 2009) revolutionized map matching by bolstering the handling of noisy GPS data and integrating contextual factors like road topology and historical trajectory patterns. Despite significant advancements, prior applications of HMM still fall short in adequately addressing the diverse range of LBS data collected from various scenarios and devices. Furthermore, there remains a lack of comprehensive research on conducting comparative analyses of different constraints and optimizing parameter tuning. Meanwhile, recent advancements in deep learning for map matching (Feng et al., 2020; Jin et al., 2022) show promise but grapple with challenges such as heavy data dependency, high computational costs, and limited interpretability, thereby impeding their application in large-scale, uncertain LBS data scenarios.

This paper aims to fill the existing gaps in LBS data by investigating adjustment methods for the HMM model and parameters, with the goal of enhancing map matching in diverse conditions. More specifically, we explore the incorporation and comparison of four supplementary constraints: Transition Detour Penalties, Transition Mobile Trends Determination, Transition Dummy Speed Determination, and Transition Link Type Determination. Additionally, we examine the parameter tuning approach using Bayesian optimization to improve map matching performance. Demonstrating effectiveness of these optimization methods based on HMM in handling LBS data, we showcase different supplementary constraints and parameter fine-tuning ability to determine optimal parameter combinations for different road networks. Our preliminary results demonstrate that the Transition Detour Penalties, Transition Mobile Trends Determination can significantly improve and enhance the performance without parameter tuning and adding the rest of the redundant conditions, giving better results than the baseline algorithms. Meanwhile, the map matching algorithm based on HMM model maintains robustness across varying data frequencies and road configurations in real LBS data.

2 METHODOLOGY

2.1 Measurement and transition probability

The first is using measurement probability to evaluate the probability of set of candidate road links $L_r = \{c_t^1, c_t^2, \cdots, c_t^k\}$ within radius r of observed points p_t ($1 \le t \le n$) in the given trajectory $T = p_1 \to p_2 \to \cdots \to p_n$. To obtain candidate road links near each observation point, the K-D Tree is used for candidate searching. Then, for each of the candidate road links, we evaluate it using measurement probability $P_M(p_t \mid c_t^k)$, Eq. 1, where δ_p is the parameter.

$$P_M(p_t \mid c_t^k) = \frac{1}{\delta_p \sqrt{2\pi}} e^{\frac{-\left\| p_t - c_{t,i}^k \right\|^2}{2\delta_p^2}} \tag{1}$$

The transition probability gives the probability of trajectory moving from one road link to the next (e.g., $c_{t-1}^k \to c_t^s$), given the observation of two location points: $p_{t-1} \to p_t$. To find the most likelihood path between two links, using bidirectional search of Dijkstra's algorithm for shortest paths. The difference between points distance $\|p_{t-1} - p_t\|$ and path length $\|c_{t-1,i}^k - c_{t,i}^s\|$ are used to calculate the transition probability $P_T\left(p_{t-1} \to p_t \mid c_{t-1,i}^k \to c_{t,i}^s\right)$ (Equation 2) from candidate road link of last point to next point.

$$P_T\left(p_{t-1} \to p_t \mid c_{t-1,i}^k \to c_{t,i}^s\right) = \frac{1}{\beta_p} e^{-\frac{1}{\beta_p}\left\| \|p_{t-1} - p_t\| - \|c_{t-1,i}^k - c_{t,i}^s\| \right\|} \tag{2}$$

Where the β_p indicate the parameter in transition probability.

2.2 Heterogeneous network and travel conditions specified in HMM

Heterogeneous network and travel conditions of different trajectories may cause errors when applied to map matching model, especially since the location data were collected without velocity or direction information and in different quality. Studies show that the HMM algorithm can be improved by adding the velocity or direction information (Hsueh and Chen, 2018; Lou et al., 2009). Optimizing the utilization of map and track data while minimizing superfluous information is a highly efficient and cost-effective strategy. In this study, to mitigate the problem of erroneous matching in the HMM algorithm, four additional conditions are introduced into HMM, making it heterogeneous and thus capable of capturing diverse situations for travel trajectories.

1. Transition Detour Penalties: Implementing stricter penalties for high-cost detours can effectively mitigate the issue of unnecessary false detours. To restrict transfers between front and back candidate roads, consider incorporating a detour penalty into the transfer probability. Equation 3 introduces μ_s as the parameter governing the penalty limit:

$$s_{di}\left(p_{t-1} \to p_t \mid c_{t-1,i}^k \to c_{t,i}^s\right) = \frac{\left\| p_{t-1}^k - p_t^s \right\|}{\left\| p_{t-1} - p_t \right\|} \le \mu_S \tag{3}$$

2. Transition Mobile Trends Determination: The mobile trend analysis incorporates changes between two points in two-dimensional space, getting rid of the single transfer calculation based on the original transfer probability which is mainly based on distance. Given the absence of angle data in numerous LBS datasets, our algorithm addresses this limitation by representing the angle between front and back points within a two-dimensional plane. Each point (except the last point) in a trajectory is assigned a dynamic trend, depicting the angular change between the two points in the plane. Subsequently, these mobile trends are compared with candidate links shifting trends, as described in Equation 4, and the new transition probability in Equation 5:

$$s_{\theta}\left(p_{t-1} \rightarrow p_t \mid c_{t-1,i}^{k} \rightarrow c_{t,i}^{s}\right) = \left|\Delta\theta_{c_{t-1}^{k} \rightarrow c_t^{s}} - \alpha_{\theta} \cdot \Delta\theta_{p_{t-1} \rightarrow p_t}\right| \tag{4}$$

3. Transition Dummy Speed determination: To address the absence of instantaneous velocity data in LBS datasets, our model incorporates dummy velocities, thereby mitigating the issue of lacking velocity, and considers oscillation which is commonly observed in LBS data. The v_{link} is speed limitation, γ ·and δ are parameters, L_T is length of a trajectory, Δt_T is the time of a trajectory.

$$s_{\bar{v}}\left(p_{t-1} \rightarrow p_t \mid c_{t-1,i}^{k} \rightarrow c_{t,i}^{k}\right) = \Delta t_T \frac{\delta \left\| p_{t-1} - p_t \right\|}{L_T} \leq \gamma \cdot v_{link} \tag{5}$$

4. Transition Link Type Determination: As the Hidden Markov process solely focuses on the current state, it tends to diminish model performance in intricate roadway scenarios, like parallel highways and high-speed intersections.

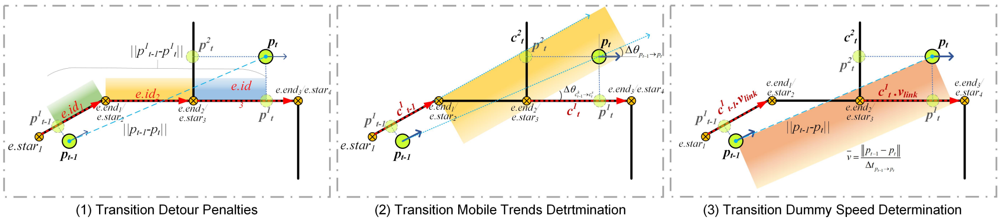

Figure 1 – *Supplement conditions* description

2.3 Parameter fine-tuning

There will be differences in the selection of parameters in HMM and supplement conditions. Bayesian Optimization (BO) effectively manages the parameter fine-tuning by selecting the most promising parameter values to evaluate, thus minimizing the number of evaluations needed. This is invaluable for many large application scenarios with location data. Furthermore, BO is a global optimization technique that can avoid local optima. This is particularly beneficial for HMM, where the parameter space can be complex and non-convex. Meanwhile, the probabilistic model maintains a belief about the state of the objective function across the parameter space, which helps in exploring the space more thoroughly and efficiently.

3 EXPERIMENTS AND RESULTS

3.1 Dataset and Evaluation metrics

To verify the model performance in varies supplement conditions, we utilize the road networks from Open Street Map (OSM) and 30 trajectories from LBS data in Washington DC area including six city/county. The LBS data includes the timestamp, latitude and latitude information, each of the trajectory with manually labeled labels for evaluation. To verify the model performance, the open-source vehicle GPS data in Seattle city (Newson and Krumm, 2009) is selected as the test dataset.

The Accuracy, Precision, Error, Recall, and F1-measure (Cui et al., 2021; Liao, 2023) are adopted as evaluation metrics. *Accuracy* is obtained by dividing number of correctly map-matched links by number of ground truth links. *Precision* is obtained by dividing length of correctly map-matched links by length of the map-matched links. *Recall* is used to measure the ability of the algorithm to correctly predict samples, which is equal to the true positive to true positive plus false negative (the total length of the ground truth trajectory); *Error* which means the erroneously subtracted (d_m^-) and added (d_m^+) among the map-matched links. The *F1-measure* is between 0 and 1 to more comprehensive evaluate of the model. The model will perform better with a larger *F1-measure*.

3.2 Results of Numerical Examples

In this sub-section, performance of benchmark HMM model and the HMM models with Bayesian optimization and different supplement constrains added were visually demonstrated, while evaluated qualitatively and quantitatively on both Dataset 1 (a vehicle GPS dataset) and Dataset 2 (an LBS-focused data set, more details will be included in the full paper). Afterwards, the improvement effect of BO on these different models was evaluated using Dataset 1 and Dataset 2.

Table 2 – *Model performance comparison in Dataset 1 and Dataset 2*

Model	*Accuracy*	*Precision*	*Recall*	*Error*	*F1*
Benchmark HMM	0.8863	0.5683	0.8794	0.6955	0.6905
HMM Transition Detour Penalties	0.8338	0.6106	0.8780	0.6429	0.7203
HMM Transition Mobile Trends	0.9928	0.6737	0.9965	0.3425	0.8039
HMM both Detour and Mobile Trends	0.9460	0.8795	0.9578	0.1662	0.9170
Benchmark HMM	0.8480	0.7110	0.8217	0.7869	0.7623
HMM Transition Detour Penalties	0.8710	0.7859	0.8503	0.5820	0.8159
HMM Transition Mobile Trends	0.8836	0.7688	0.8481	0.6165	0.8065
HMM both Detour and Mobile Trends	0.8947	0.8037	0.8483	0.3833	0.8254

4 Conclusion and Future Work

The study develops a novel Bayesian-optimized Hidden Markov model to recognize network paths using location-based service data trajectories. The proposed approach has been demonstrated via a numerical test of two sample datasets, showing a significant effect of the complementary conditions on the improvement of map matching. Our results surpass the F1-measure in a GPS-based Dataset 1 reported in earlier studies (i.e., Nearest Neighbor: 0.703; Hybrid Model: 0.913; Liao, 2023). The performance of such approach on an LBS Dataset 2 was also improved from its benchmark (to an overall F1 score of 0.8254). However, clearly the performance was worse than Dataset 1 models. This is likely due to the high fluctuation of LBS datasets in terms of the location oscillation and unstable observation frequency. The model development of adding speed and link types as prior information is currently under testing and anticipated to further improve the model performance.

References

Cui, G., Bian, W., Wang, X., 2021. Hidden Markov map matching based on trajectory segmentation with heading homogeneity. *GeoInformatica* 25, pp. 179–206.

Feng, J., Li, Y., Zhao, K., Xu, Z., Xia, T., Zhang, J., Jin, D., 2020. DeepMM: Deep Learning Based Map Matching with Data Augmentation. *IEEE Trans. Mob. Comput.* pp. 1–1.

Hsueh, Y.-L., Chen, H.-C., 2018. Map matching for low-sampling-rate GPS trajectories by exploring real-time moving directions. *Inf. Sci.* 433–434, pp. 55–69.

Jin, Z., Kim, J., Yeo, H., Choi, S., 2022. Transformer-based map-matching model with limited labeled data using transfer-learning approach. *Transp. Res. Part C Emerg. Technol*. 140.

Liao, J., 2023. Optimization of Map Matching Algorithm in Various Road Conditions. *Highlights Sci. Eng. Technol.* 78, 59–66.

Lou, Y., Zhang, C., Zheng, Y., Xie, X., Wang, W., Huang, Y., 2009. Map-matching for low-sampling-rate GPS trajectories, *17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems,* ACM, Seattle Washington, pp. 352–361.

Newson, P., Krumm, J., 2009. Hidden Markov map matching through noise and sparseness, *17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ACM, Seattle Washington, pp. 336–343.