EXPLORING URBAN MOVEMENTS: A DATA-DRIVEN PERSPECTIVE ON HUMAN MOBILITY MINING

Shafqat Ali Shad[1]*, Muhammad Usman.[2], Chandan Kumar[3], Hadiqa Afzal[4]
University of Montevallo, Montevallo, Alabama, USA[1,4]
Iowa State University, Iowa, USA[3]
Verizon, Florida, USA[2]
Email: sshad@montevallo.edu, m.usman30@outlook.com, chandan@iastate.edu,, hafzal@montevallo.edu,

ABSTRACT- *The exponential growth of location acquisition technologies has created a rich landscape of spatiotemporal data, offering unparalleled opportunities to understand human movement and advance location-based services (LBS). This paper outlines a methodology to derive user trajectories and analyze recurring behavioral patterns from GPS log data. A core proposition tested is the observation that individuals tend to spend a significant majority (approximately 80%) of their time within familiar or "safe" environments. To substantiate this, raw GPS patterns were transformed into records of visited locations, which were then analyzed for daily and periodic trends over extended periods. The study utilized GPS data from Microsoft Research Asia, comprising records for 182 users, with detailed movement mode information available for 73 participants.*

Keywords: Human movement, spatial analysis, behavioral insights, location data, urban analytics

1. INTRODUCTION

The widespread adoption of modern location technologies, including GPS, GSM, and Wi-Fi, has led to an abundance of spatiotemporal data [1]. This data offers a powerful lens through which to examine human mobility, revealing intricate details about individual trajectories, daily behaviors, ingrained habits, and summaries of visited places. Such insights are fundamental for developing a new generation of location-based services. The applications of understanding these mobility patterns are vast, ranging from developing proactive early warning systems and predicting travel routes to informing targeted marketing campaigns, optimizing urban transportation, forecasting future locations, unearthing deeper habits, and even constructing social networks [2].

Raw mobility data, despite its richness, often requires transformation into more meaningful representations to be directly applicable to specific analytical challenges. Prior research has explored various aspects of this domain. For example, [3] focused on predicting users' next locations, while [4] investigated its use in building social networks. [5] innovatively combined mobility data with semantic or contextual tags to gain insights into typical user behaviors. Similarly, [6] concentrated on converting raw mobility data into geocodes or "stay points" that could subsequently be analyzed for user habits and trends.

However, the precise identification of stay points and geocodes directly from raw mobility data presents a significant analytical hurdle. Over the years, researchers have proposed numerous techniques to derive meaningful life trends or conduct habit mining from this complex data. Recognizing that raw mobility data alone is often insufficient for developing compelling real-world applications, this study proposes a novel approach for extracting stay points from raw GPS data and analyzing user habits by examining both place exploration and visit summaries. This effort is critical for advancing our understanding of human mobility and its profound implications for sophisticated location-based services [7]. By employing advanced spatiotemporal data mining techniques, our methodology offers a systematic framework for uncovering latent patterns and trends within mobility data, thereby contributing significantly to both foundational research and practical advancements in this evolving field.

2. LITERATURE REVIEW

Previous research has significantly contributed to various dimensions of human mobility analysis. [8] pioneered the use of semantic tags to identify significant locations and extract visit histories, subsequently proposing a combination of semantic and geotags for future visit recommendations. Their work also touched upon urban sensing through crowdsourcing to identify city hotspots, a concept discussed by [9]. [10] investigated user similarity based on movement trajectories, incorporating semantic information into their analysis. While some studies, such as [11], attempted to determine transportation modes from GPS trajectories, they frequently overlooked broader user patterns or historical contexts of movement.

The concept of location history has gained increasing prominence in various recommender systems, including platforms like [12] and CityVoyager, which recommend popular urban locations such as shops and restaurants. [13] proposed the application of collaborative filtering for restaurant recommendations. Similarly, [10] and [14] aimed to leverage user history and visitation trends for shop recommendations; however, these approaches often relied on traditional recommender system technologies rather than integrating true spatiotemporal trends, the sequential nature of user movements, and geospatial hierarchies. More recently, there has been a growing interest in using human mobility data to generate synthetic social networks, with [15] exploring the synthesis of social networks based on human mobility patterns to understand the interplay between movement and social interactions.

Building upon this existing foundation, our study delves into the intricacies of human mobility by directly utilizing raw GPS data to uncover patterns and behaviors. Unlike prior works that often heavily relied on semantic and geotags to discern significant places and recommend subsequent visits, our approach emphasizes the analysis of raw GPS data to derive detailed user trajectories and movement histories. We introduce specific distance and time threshold values for precisely identifying "stay points" or significant locations, demonstrating their effectiveness through empirical validation.

While we explored the integration of semantic information extracted via the Google Geolocation API for topic modeling, practical challenges arose due to the dataset predominantly featuring users from Beijing, leading to difficulties in processing Chinese character addresses [16].

Our research embarks on a new direction by directly exploring spatiotemporal trends through raw GPS data, diverging from the conventional reliance on semantics and geotags. This direct analysis of user trajectories promises a more authentic and granular understanding of individual movements through space and time. The rising popularity of location history-driven recommender systems underscores the necessity of incorporating genuine spatiotemporal insights into recommendation algorithms, as traditional methods often fall short in capturing the subtle preferences of users [17]. Our work aims to address this by leveraging GPS data to offer personalized recommendations meticulously tailored to each user's unique trajectories, thereby enhancing both relevance and precision. Through rigorous experimentation and validation, we demonstrate the efficacy of our approach in unraveling complex human movement patterns [18]. This study represents a significant advancement in decoding the intricate tapestry of spatiotemporal trends solely through GPS data analysis. By shedding light on the fundamental complexities of human mobility, we lay the groundwork for developing more sophisticated and context-aware location-based services, highlighting the transformative potential of advanced methodologies to interpret human movement within urban environments [19].

3. METHODOLOGY

Our methodology primarily harnesses GPS log data, specifically latitude and longitude coordinates, to meticulously reconstruct individual user trajectories. These reconstructed trajectories, in turn, serve as the basis for defining unique user-specific mobility trends, which collectively reveal broader community-level movement patterns [20]. Before detailing our approach, it is essential to define the key terms used throughout this study:

GPS log (G): Represents the complete collection of all GPS records for a given user *U*, formally denoted as $G=\{g_1,g_2,g_3,...,g_n\}$.

Stay point or significant place (SP): A distinct set of GPS points, $SP=\{sp_1,sp_2,sp_3,...,sp_n\}$, identified by applying specific time and distance thresholds (TH_{tm} and TH_{ht}). Stay points signify locations where a user has spent a notable amount of time, as determined by these thresholds [21].

Unique location (UL): Refers to the set of distinct locations, $UL=\{ul_1,ul_2,ul_3,...,ul_n\}$, identified throughout the mobility history of user *U* over a defined period *T*.

Patterns and trajectory (T): Describes sequential user movements from one unique location to another within a specified day, represented as $T=\{t_1,t_2,t_3,...,t_n\}$, and must meet a minimum number of locations threshold ($TH_{loc_{num}}$).

Frequent patterns (FP): These are patterns or trajectories that recur with a frequency meeting or exceeding a minimum frequency threshold ($TH_{freq_{min}}$) and incorporate a decay value (*DV*). These are crucial for identifying persistent movement trends and recurring behavioral patterns [22].

Phase 1: GPS Data Extraction

The initial phase involves extracting the necessary GPS information from the designated dataset. The GPS dataset is typically provided in the Hewlett-Packard Graphics Language (HPGL) plot (PLT) format, which is easily parsable and contains daily GPS logs for each user. This raw GPS data constitutes the fundamental input for all subsequent trajectory analysis and pattern recognition algorithms.

Phase 2: Stay Point Identification using Temporal and Spatial Thresholds

For the precise identification of stay points, we implemented an algorithm (refer to Algorithm 1 in the original document) with an efficient time complexity. This algorithm judiciously applies a time threshold (THtm) of 30 minutes and a distance threshold (THdt) of one mile to the GPS data extracted in Phase 1. This method systematically identifies periods and locations where a user's movement suggests a deliberate stop rather than transit.

Phase 3: Deriving Unique Locations from User Historical Mobility Data

The data from Phase 2, specifically the extracted stay points, was then used to identify unique locations from each user's complete mobility history. A distance threshold (THdt) of one mile was consistently applied to aggregate geographically close stay points into single unique locations. This threshold was chosen based on an analysis of typical human movement characteristics, considering the interplay between distance, time spent, and location proximity.

Phase 4: Extraction of Mobility Patterns Over Unique Locations

To comprehensively extract mobility trajectories and patterns, we leveraged the unique location information obtained from Phase 3 and applied a dedicated algorithm (refer to Algorithm 2 in the original document). This process involves converting unique location sequences into daily records, applying an association rule mining algorithm (like APRIORI) to find frequent sequences, and then filtering these to retain only recent frequent patterns using a decay value to manage data volume.

3. DATA SPECIFICTAION

The public dataset utilized in this study originates from the Geolife project, an initiative by Microsoft Research Asia in 2020. This extensive repository of GPS logs covers a three-year period and documents the mobility patterns of 182 individual users. Each entry in the dataset provides rich temporal and spatial information, including precise location coordinates and corresponding timestamps. Furthermore, for 73 users, the dataset includes additional insights into their mode of movement, such as travel by bus, train, walking, or airplane. It is noteworthy that the majority of this data pertains to users located within Mainland China, offering a unique perspective on urban mobility dynamics in that region.

The Geolife dataset offers a multifaceted view of human mobility, enabling detailed analyses of movement patterns and behaviors over an extended temporal horizon. Its granular temporal resolution allows researchers to delve into the minutiae of daily routines and travel trajectories, illuminating the complexities of urban mobility dynamics. The inclusion of movement mode data further enriches the dataset by providing insights into transportation preferences and shifts across various contexts. Consequently, the Geolife dataset serves as an invaluable resource for studying human mobility at scale, facilitating the discovery of spatiotemporal trends, the discernment of habitual behaviors, and the development of innovative solutions for location-based services and urban planning.

4. RESULTS

Our experimental phase involved a series of analyses aimed at validating the proposed methodology and gaining a deeper understanding of various aspects of human mobility.

Experiment 1: Optimizing Stay Point Extraction

A critical aspect of our methodology was the careful determination of the time threshold (THtm). We conducted a thorough investigation of various time thresholds, including 10, 20, 30, and 60 minutes, to assess their impact on the accurate identification of unique locations. By systematically varying the time threshold, we sought to understand how different durations influence the detection of stay points and the overall characterization of human mobility patterns. Our findings indicated that a 30-minute time threshold yielded an optimal number of stay points, striking an effective balance between capturing fine-grained detail and ensuring comprehensive coverage of significant locations. This moderate threshold helped avoid both over-segmentation (too many stay points) and under-segmentation (missing important locations), ensuring accurate capture of human mobility while minimizing noise from raw GPS data. Figure 4 (in the original document) illustrates this impact for a random user.

Experiment 2: Characterizing Unique Historical Locations

Following the establishment of the distance threshold (THdt), we conducted a comprehensive analysis of the entire dataset, which spans three years of GPS data for each user. This holistic approach allowed us to capture the full spectrum of an individual's mobility patterns over an extended period and identify unique locations that persisted over time, providing insights into habitual user movements and favored destinations. A consistent distance threshold of one mile was applied to ensure that significant locations were consolidated while filtering out minor GPS fluctuations.

We observed the effectiveness of this distance threshold in delineating unique locations for a random user, successfully transforming raw GPS data into meaningful clusters corresponding to distinct visited locations over the three-year period. These unique locations serve as nodes in the user's mobility network, each characterized by its visit frequency and temporal patterns. Quantifying and visualizing these unique locations offered valuable insights into user travel behavior and preferences, forming the foundation for further analysis. This highlights the potential for integrating this threshold into broader mobility analytics frameworks, enhancing the ability to interpret human movement patterns across diverse contexts and facilitating comparisons between datasets. Figure 5 (in the original document) depicts unique location data for 10 random users.

From the extracted unique location information across multiple users, we found that the number of unique user locations ranged from 30 to 700, with an average of 200 unique locations over three years of geolocation history. This led us to conclude that human life trends suggest individuals spend approximately 80% of their lives in known places, rather than constantly exploring new ones, unless a relocation occurs.

Uncovering Mobility Trajectories and Patterns

To identify frequent mobility patterns over time, we introduced a "stay time" (Tstay) of 15 minutes and a "transition time" (Ttran) with a distance threshold of one mile. Our analysis revealed that a sample user, on average, exhibited approximately 173 unique patterns. The variability was notable, with some users demonstrating a maximum of 350 unique patterns and others showing a minimum of seven patterns. This variability underscores the diverse nature of human mobility behaviors and the importance of considering individual preferences and routines.

The inclusion of transition time enabled us to capture not only the spatial aspect of mobility but also the temporal dynamics involved in moving between different locations. This temporal dimension facilitated the identification of recurring movement patterns that manifest over time, offering insights into the regularity and predictability of an individual's travel behavior. By quantifying the frequency of unique patterns for each user, we gained a deeper understanding of the underlying structure of human mobility and the factors driving variations across different individuals and contexts.

This temporal view allowed us to identify recurring movement patterns over time, providing insights into the regularity and predictability of an individual’s travel behavior. Quantifying the frequency of unique patterns for each user deepened our understanding of human mobility's structure and the factors influencing variations across individuals. Our findings reinforce that the choice of transition time and distance threshold significantly affects the granularity and comprehensiveness of extracted patterns. This iterative approach underscores the importance of parameter tuning in data-driven analyses, highlighting the value of considering spatial proximity and temporal continuity in delineating meaningful patterns, crucial for urban planning, transportation, and LBS.

Most users exhibit fewer than 100 patterns, but none fewer than five, over the data recording period. This variability highlights the rich diversity of human life and activity, emphasizing the complex interplay between individual preferences, environmental factors, and social dynamics. Understanding these patterns is essential for gaining insights into human behavior and developing effective strategies for urban planning, transportation management, and community development.

When trajectory direction or the sequence of activities is considered, the size of the database shrinks, leading to a more focused analysis compared to non-sequenced patterns. This sequential representation allows for a more granular examination of individual behaviors, enabling the discovery of underlying patterns not apparent in aggregate data. It also opens opportunities for predictive modeling and recommendation systems by capturing temporal dynamics and identifying recurring activity sequences. Moreover, analyzing sequential movements can deepen understanding of social interactions and community dynamics, revealing gathering spots and common routes, and informing decision-making in emergency planning, urban development, and public health by correlating mobility with contextual variables.

It is evident that the number of patterns considered with sequence is much lesser compared to those without sequence. While a different sequence of visits daily is common, the underlying unique locations for a particular user remain limited, as do the mobility patterns over these unique locations.

5. CONCLUSION

Our study successfully leveraged GPS information to extract stay points without relying on geotagging, enabling the accurate derivation of human mobility patterns. A key finding from our analysis is that users typically spend approximately 80% of their lives in known places, with only 20% dedicated to exploring new locations. This suggests that the number of unique locations frequented by individuals remains finite and limited, as do their recurring mobility patterns.

These insights open new opportunities across various fields, including social networking, community detection, and urban mobility planning. Furthermore, most mobility-based recommender systems can effectively integrate our work, adapting it to the specific nature of their recommendations.

In essence, our study illuminates the fundamental characteristics of human mobility patterns, emphasizing the prevalence of recurrent visits to familiar locations and the occasional exploration of new venues. By utilizing GPS data to extract stay points and analyze mobility behaviors, we have uncovered insights with far-reaching implications across diverse domains. From informing social networking strategies to guiding urban planning and enhancing recommender systems, our findings offer valuable guidance for designing and implementing tailored solutions that harness the predictable nature of human mobility to improve user experiences and optimize resource allocation.

Moving forward, further research in this area holds significant promise for developing advanced algorithms and methodologies capable of uncovering deeper insights into human behavior. This will facilitate the creation of intelligent systems that can adapt to the dynamic needs and preferences of individuals and communities. Our work underscores the importance of data-driven approaches for understanding human mobility dynamics. By harnessing the rich information embedded within GPS logs and other location-based datasets, researchers and practitioners can continue to unravel the complexities of human behavior, paving the way for innovative solutions to real-world challenges.

REFERENCES

1. Adams, L., Lu, Y., & Liu, Y. (2012). Pervasive location acquisition technologies: Opportunities and challenges for geospatial studies. *Computers, Environment and Urban Systems, 36*(2), 105–108.
2. Arya, A., Tiwari, V., & Arya, A. (2017). Horizontally scalable probabilistic generalized suffix tree (PGST) based route prediction using map data and GPS traces. *Journal of Big Data, 4*(1).
3. Mehdizadeh, M. S., & Bahrak, B. (2020). A regression framework for predicting user's next location using Call Detail Records. *Computer Networks, 183*, 107618.
4. Lai, S., et al. (2019). Exploring the use of mobile phone data for national migration statistics. *Palgrave Communications, 5*(1), 34.
5. Guo, L. (2012). The application of social network analysis in agenda-setting research: A methodological exploration. *Journal of Communication, 62*(4), 616–631.
6. Gao, S., Gao, H., & Liu, H. (2015). Mining human mobility in location-based social networks. *Synthesis Lectures on Data Mining and Knowledge Discovery, 7*(2), 1–115.
7. Shad, S., Usman, M., Kumar, C., & Afzal, H. (2024). Understanding places exploration and visitation via human mobility mining. *International Journal of Intelligent Information Technologies (IJIIT), 20*(1), 1–16.
8. Hamamerh, A., Hamamerh, R., & Awad, S. (2018). Intelligent social networks model based on semantic tag ranking. *International Journal of Web and Semantic Technology, 9*(3), 1–12.
9. Senanayake, J. M. D., & Wijayanayake, W. M. J. I. (2018). Applicability of crowd sourcing to determine the best transportation method by analysing user mobility. *International Journal of Data Mining and Knowledge Management Process, 8*(4/5), 27–36.
10. Basu, L., & Li, T. (2019). A machine-learning-based early warning system boosted by topological data analysis. *SSRN*.
11. Mankad, M., Verma, J., Mankad, S., & Garg, S. (2020). GeoHash tag based mobility detection and prediction for traffic management. *Springer Nature Applied Sciences, 2*(8).
12. Geowhiz. (2020). Geowhiz: We create worlds of fun! http://www.geowhiz.com
13. Füser, M. W., et al. (2018). SEDE-GPS: Socio-economic data enrichment based on GPS information. *BMC Bioinformatics, 19*(Suppl 15).
14. Liu, Y., & Liu, H. (2018). A tag-based recommender system framework for social bookmarking websites. *International Journal of Web Based Communities, 14*(3), 1.
15. Gallagher, K., et al. (2022). Human mobility-based synthetic social network generation. *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Animal Movement Ecology and Human Mobility*, 23–26.
16. Microsoft Research. (2020). Geolife: Building social networks using human location history. https://www.microsoft.com/en-us/research/project/geolife-building-social-networks-using-human-location-history
17. Khayyambashi, M. R., Movahedian, H., & Khayyambashi, M. H. (2015). A semantic recommender system based on frequent tag pattern. *Intelligent Data Analysis, 19*(1), 109–126.

18. Kumar, S. S., Sakthivel, D., & Kumar, M. P. (2016). To motivate buyers habit based on WhatsApp data using data mining techniques. *International Journal of Computer Trends and Technology, 41*(1), 26–28.
19. Szell, M. (2018). Crowdsourced quantification and visualization of urban mobility space inequality. *Urban Planning, 4*(1), 1–10.
20. Elalaouy, R., Rhoulami, K., & Driss, M. (2017). A novel modeling based agent cellular automata for advanced residential mobility applications. *International Journal of Advanced Computer Science and Applications, 8*(7).
21. Tran, D., et al. (2018). A Closer Look at Spatiotemporal Convolutions for Action Recognition. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6450–6459.
22. Dey, H., Rúa, S., & Dey, D. (2019). A Bayesian piecewise survival cure rate model for spatially clustered data. *Spatial and Spatio-Temporal Epidemiology, 29*, 149–159.