Semi-Automatic Detection and Evaluation of Cyclist Points of Interest Using Riding Log Data Analysis

Kazuki Nagayama[1], Yuki Matsuda[3,1,2], Hirohiko Suwa[1,2], Keiichi Yasumoto[1,2],
[1] Nara Institute of Science and Technology, Ikoma, Nara 630-0192, Japan
Email:nagayama.kazuki.nh3@naist.ac.jp, {yukimat, h-suwa, yasumoto}@is.naist.jp
[2] RIKEN, Center for Advanced Intelligence Project, Tokyo 103-0027, Japan
[3] Okayama University, Okayama 700-8530, Japan

***Abstract*—There are places that are popular with cyclists (cyclist points of interest: CPoI). However, these CPoIs are often only known within the cyclist community and are not widely shared as general tourist information. In this study, we propose a semi-automated method for detecting and evaluating CPoIs from cyclist ride log data. To verify the proposed method, we used 374 detailed ride data collected at "Michi-no-Eki Maebashi Akagi" in Maebashi City, Gunma Prefecture. By applying stop location estimation and hierarchical clustering to the collected data, we identified 125 CPoIs. We then used crowdsourcing to identify and evaluate the CPoIs. As a result, 109 locations (87.2%) matched our expectations, and a positive correlation was observed between the evaluation results. Furthermore, by analyzing the relationship between the evaluation of CPoIs and the number of visitors, we showed that it is possible to distinguish between popular CPoIs, which are widely recognized, and hidden CPoIs, which are highly rated but not well known. These results show that the proposed method is effective for detecting and evaluating CPoIs.**

***Index Terms*—cycle tourism, cyclist points of interest (CPoI), crowdsourcing, hierarchical clustering, tourist spot detection**

I. INTRODUCTION

Cycle tourism is attracting attention as an environmentally friendly and sustainable form of tourism. This tourism style is expected to contribute to regional revitalization as cyclists discover new local attractions while leisurely exploring peaceful landscapes and natural settings, as well as through interactions with local residents. In particular, resources often overlooked by conventional tourism—such as small-scale facilities that can be explored briefly, natural scenery, and quiet roadside views—have the potential to become attractive stopping points for cyclists (Cyclist Points of Interest: CPoI). However, these CPoIs tend to be shared only within specific cyclist communities, and efficient methods for identifying and sharing them as new tourism resources have not been established. In this study, we propose a semi-automated method for detecting and evaluating CPoIs from cyclists' ride log data. Specifically, we developed a method to identify CPoIs through stopping point estimation and clustering of cyclists' ride log data, along with a methodology to evaluate CPoI characteristics using crowdsourcing. Applying the proposed method to actual data collected in Maebashi City, Gunma Prefecture, we successfully identified 125 significant CPoIs. Furthermore, through analyzing the relationship between CPoI evaluations and visitor numbers, we gained new insights into distinguishing between widely recognized spots (Popular CPoIs) and highly-rated yet less-known spots (Hidden CPoIs) based on the correlation between visitation frequency and evaluation scores. The effectiveness of our proposed method was validated by the crowdsourcing identification results, where 87.2% of the detected CPoIs matched the author's expectations, and positive correlations were observed between evaluation results.

II. RELATED WORK

A. Analysis of Cyclist Behavior

Numerous studies have analyzed GPS data to investigate cyclists' route choice behaviors [1]–[7]. In the context of tourism and GPS studies, Ritchie et al. [1] revealed cyclists' characteristics and infrastructure usage patterns in tourism contexts, while Prato et al. [2] proposed a bicycle route choice model in distance-value space using GPS data from the Copenhagen region. In research utilizing bike-sharing system data, Khatri et al. [3] and Scott et al. [4] analyzed user behavior patterns and preferences in Phoenix and Hamilton, respectively. Chen et al. [5] examined how land use and road network characteristics influence cyclists' route preferences, while Zimmerman et al. [6] proposed an efficient route choice model using a recursive logit model. These studies have revealed clear patterns in cyclists' location preferences. Cyclists show preferences for riding through residential areas and scenic locations [2], [3], and tend to favor quick travel on low-traffic roads and dedicated cycling infrastructure [3], [5]. Some cyclists also show preferences for routes with street trees, street lighting, and routes surrounded by mixed land use [5]. Conversely, cyclists tend to avoid unpaved roads, hilly terrain, and cycling along major roads [2], [5]. Additionally, detour routes, corners, steep slopes, high-traffic roads, and one-way sections have been shown to negatively impact route selection [3], [4].

B. Extraction of PoIs from Location Data

Numerous studies have focused on extracting Points of Interest (POIs) from location data [8]–[12]. In research utilizing location-based User Generated Content (UGC) such as Flickr, Kuo et al. [8] proposed an efficient method for extracting POIs and Regions of Interest (ROIs) from geotagged photos, while Kozaki et al. [9] conducted research on discovering hotspots from social media data. In studies identifying POIs from GPS data, Krause et al. [10] proposed an automated activity location identification method combining GPS and land use data for short-term movement prediction, while Ashbrook et al. [11] developed a Markov model for learning significant locations and predicting movement patterns across multiple users. Additionally, Zheng et al. [12] proposed a method for estimating users' interest levels in locations based on personal location histories. Furthermore, substantial research has been conducted on discovering lesser-known Hidden POIs [13]–[16]. Kitayama et al. [13] proposed a method for extracting attractive hidden spots based on user evaluations and visitor numbers, while Zhuang et al. [14] developed a method for discovering hidden tourist spots by evaluating popularity and scenic quality based on photographer behavior analysis. Regarding POI evaluation, Yang et al. [17] proposed a method for calculating recognition and user evaluations using social media data, and Cui et al. [18] developed a methodology for ranking Hidden POIs using crowdsourcing.

C. Identification of Stay Points Using GPS Data

Recent years have seen active research in identifying stay points and analyzing activities using GPS data [19]–[24]. In methods utilizing POI data, Liu et al. [19] developed an automated approach for identifying urban blocks using OpenStreetMap and POI data, while Huang et al. [20] proposed a method for identifying activity locations from GPS data by defining spatial and temporal attractiveness of POIs. Additionally, Zhao et al. [21] achieved high-accuracy travel purpose prediction by combining GPS data and POI information using gradient boosting decision trees. For more advanced analytical approaches, Lyu et al. [22] developed a destination prediction model based on understanding travel purposes by combining GPS and land use data. Furthermore, Tao et al. [23] achieved high-accuracy activity type detection (96.8%) using random forests by integrating spatial and temporal information.

III. Proposed Method

In this study, we propose a method for semi-automatically detecting and evaluating CPoIs using cyclists' ride log data. Figure 1 shows an overview of the proposed method. The cyclists' stopping points are estimated from ride log data, and these locations are considered CPoI candidates. CPoIs are detected by applying hierarchical clustering to these candidates. Crowdsourcing is used to identify buildings and facilities at the detected CPoI locations and to evaluate the attractiveness of CPoIs.

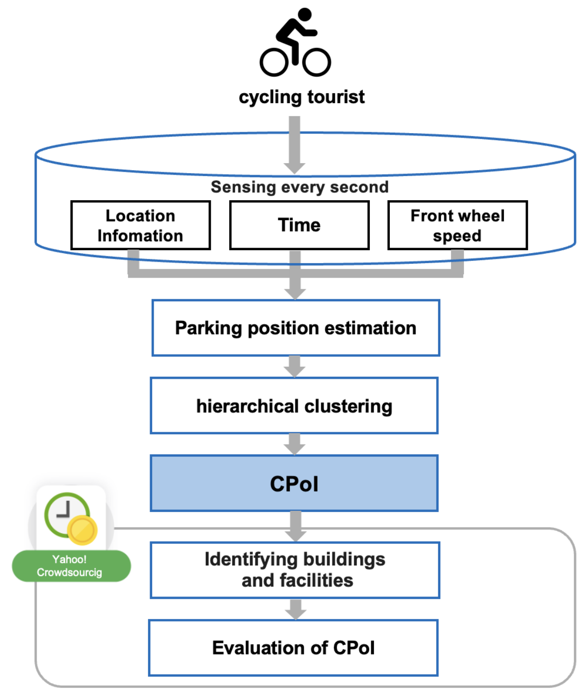

Fig. 1. Overview of the Proposed Method

A. Structure of Ride Log Data

Ride log data is recorded as time-series tracking data of each cyclist's movement. This data includes the cyclist's position information (latitude and longitude), bicycle speed information, and time information at one-second intervals.

B. Estimation of Stopping Points

To extract meaningful CPoIs, we first need to narrow down significant points from the ride log data. Since CPoIs are locations for rest, sightseeing, and enjoying scenery, it is appropriate to identify cyclists' stay points. Stay points are defined as GPS coordinates recorded when the front wheel speed remains below a threshold v for t seconds or longer. In our experiment, we set the threshold v to $5m/s$ and t to 180 seconds. The value of v was set as a speed that clearly indicates stopping, considering typical bicycle riding speeds and GPS errors. The value of t was determined to be 3 minutes as the appropriate threshold because durations less than 3 minutes might include temporary stopping points such as traffic signals and route checking, while durations longer than 3 minutes might miss short rest stops or scenic viewing locations.

C. Location Clustering

Due to GPS measurement errors of several meters, the same location may be recorded with different coordinates. Therefore, we determine CPoIs by clustering locations. For this purpose, we use Ward's minimum variance method [25], which is one of the hierarchical clustering algorithms. Ward's minimum variance method minimizes the increase in variance when merging clusters. Specifically, we first calculate the Euclidean distances between each stopping point to create a distance matrix. Next, we sequentially merge the closest stopping points or clusters, forming clusters that minimize the increase in within-cluster variance during this process. Finally, we stop merging when the increase in variance during cluster combination exceeds a threshold, thus identifying the final CPoIs.

D. CPoI Identification Using Crowdsourcing

After narrowing down meaningful CPoIs from a large amount of ride log data, we identify what buildings and facilities exist in each cluster. When there are multiple location candidates, we need to select the ones that cyclists would particularly prefer. In our method, we use Yahoo Crowdsourcing[1] for location identification. Crowdsourcing excels in providing objective evaluations from numerous participants at low cost and efficiently collecting geographical information judgments [26], [27]. The specific identification procedure is as follows:

1) Calculate the center coordinates of each cluster.
2) Obtain Street View images and surrounding maps of the center coordinates.
3) Use Google Maps Places API[2] to obtain a list of nearby buildings and facilities.
4) Through the crowdsourcing platform, present participants with Street View images and surrounding maps, and ask them to select the buildings and facilities visible in the images from the list obtained via Google Maps Places API.

The list of buildings and facilities is limited to the five closest locations within a 100-meter radius from the center coordinates. This consideration accounts for cases where bicycle parking locations might be separate from the intended destination, and the fact that having more than five options could complicate decision-making and potentially lead to less accurate responses.

E. CPoI Evaluation Using Crowdsourcing

While it is necessary to evaluate whether the identified CPoIs are actually attractive to cyclists, conventional methods are inadequate due to the lack of traditional evaluation metrics such as check-in records and reviews. Therefore, in this study, we employ crowdsourcing to evaluate CPoIs. Crowdsourcing is widely used for evaluating hidden spots with limited data [18] and efficiently collecting human judgments on geographical information from images [26], [27]. We consider this method appropriate for our study, which requires objective evaluations from numerous users. Specifically, we present participants with maps and Street View images centered on the locations identified in Section III-D and ask them to evaluate them using a four-point Likert scale based on criteria important to cyclists (Table I), such as scenic quality, suitability as a rest stop, appropriateness as a dining location, and appeal as a tourist spot. If the Street View information is outdated and the identified building cannot be confirmed, participants are asked to select the option "specified building cannot be found," and the average of valid evaluations is used as the attractiveness score for that location.

[1]https://crowdsourcing.yahoo.co.jp/

[2]https://developers.google.com/

TABLE I
EVALUATION CRITERIA FOR CPOIS

Rating	Criterion: Suitability as a CPoI
4	Highly suitable
3	Moderately suitable
2	Moderately unsuitable
1	Highly unsuitable
-	Cannot find the specified building/facility

IV. EXPERIMENT

A. Data Collection

In this experiment, the ride log data is collected from cyclists using rental bicycles at Michi-no-Eki Maebashi Akagi in Maebashi City, Gunma Prefecture, in collaboration with E-Force Inc.[3]. Data collection was conducted over a nine-month period from April to December 2023, obtaining detailed riding data at one-second intervals from a total of 374 participants.

B. CPoI Detection Results

In this experiment, we analyzed cyclists' ride log data to detect CPoIs. First, we extracted stopping points where speed remained below 5m/s for more than 180 seconds, identifying 896 initial CPoI candidates. As GPS measurements may record the same location with different coordinates due to measurement errors, we applied Ward's minimum variance method for hierarchical clustering to these candidates. Ward's minimum variance method allows adjustment of cluster numbers by changing the threshold of variance increase rate. Too many clusters lead to detection of duplicate similar locations, while too few clusters result in locations with different characteristics being treated as the same CPoI. Additionally, appropriate threshold values depend on regional characteristics such as tourist spot density and topographical features. In this study, we empirically determined the number of clusters to be 125 after testing different threshold values. The 125 detected CPoIs were distributed across urban areas, along major roads, and in suburban and natural areas. Figure 2 shows the spatial distribution of these CPoIs on a wide-area map.

C. CPoI Identification Results Using Crowdsourcing

We utilized crowdsourcing to identify the 125 CPoIs detected through hierarchical clustering. As shown in Figure 3, we presented each CPoI to 10 participants with Street View images and surrounding maps centered on its location, asking them to select appropriate places from candidates within a 100-meter radius obtained through Google Maps Places API. As a result of the identification process, we were able to identify specific place names for 79 locations (63.2%). These locations encompassed a wide range, including tourist spots such as local landmarks and historical buildings, dining facilities like cafes, restaurants, and roadside stations, public facilities such as parks, observation decks, and train stations, and commercial facilities including convenience stores and bicycle shops. For the remaining 46 locations (36.8%), most were confirmed to be roadside points without specific buildings or facilities nearby.

[3]https://www.eforce.co.jp/

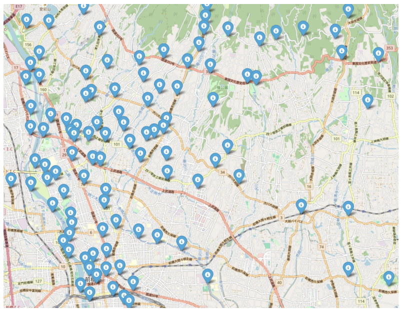

Fig. 2. Distribution Map of Detected CPoIs

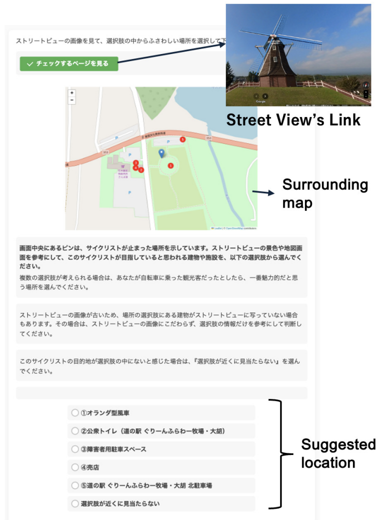

Fig. 3. Sample Crowdsourcing Interface for CPoI Identification

Fig. 4. Sample Crowdsourcing Interface for CPoI Evaluation

D. CPoI Evaluation Results Using Crowdsourcing

We conducted CPoI evaluations using crowdsourcing. As shown in Figure 4, ten participants per CPoI were presented with Street View images and asked to rate the location's attractiveness to cyclists on a four-point scale, with the average score serving as the evaluation value for that CPoI. Four locations where the response "Cannot find the specified building/facility" exceeded half of the responses were excluded from the evaluation, as these were locations where buildings could not be confirmed due to outdated Street View information. Table II shows the top 10 highest-rated CPoIs. High ratings were obtained by tourist facilities such as the Dutch-style Windmill and Santai Shrine, public facilities including Asahigaoka Park and Chuo-Maebashi Station, agricultural facilities like Tokunaga Farm and Fukubuta-no-sato Ton-ton Square, and rest facilities such as Isshin-an, Himeyuri Parking Lot, FamilyMart Shibukawa Bando Bridge Store, and 7-Eleven Maebashi Komagata Inter Store.

TABLE II
TOP 10 HIGHEST RATED CPOIS

Rank	Place Name	Rate
1	Dutch-style Windmill	3.89
2	Asahigaoka Park	3.89
3	Santai Shrine	3.88
4	Chuo-Maebashi Station	3.88
5	Tokunaga Farm	3.86
6	Isshin-an	3.78
7	Himeyuri Parking Lot	3.71
8	Fukubuta-no-sato Ton-ton Square	3.71
9	FamilyMart Shibukawa Bando Bridge Store	3.71
10	7-Eleven Maebashi Komagata Inter Store	3.70

V. DISCUSSION

A. Evaluation of CPoI Identification Results Using Crowdsourcing

In this study, we used crowdsourcing to identify CPoIs. To validate this method's effectiveness, we examined each CPoI stopping position data, predicted cyclists' destinations, and compared these with the locations determined through crowdsourcing. The verification revealed that 109 out of 125 CPoIs matched our predicted destinations, achieving an accuracy of 87.2% (109/125). This result demonstrates that our proposed method can identify CPoIs with relatively high accuracy. Analysis of the identified CPoIs indicates that cyclists tend to prioritize factors such as scenic quality, suitability for breaks, and tourist appeal. Notably, locations combining multiple of these elements received higher evaluations. However, most identification failures stemmed from discrepancies between destinations and parking locations. For example, at mountain shrines, multiple stopping points were detected both around the shrine itself and at parking areas at the mountain base. When these were processed as a single cluster, the cluster center shifted toward the base, preventing accurate identification. This issue was particularly prominent in tourist spots with elevation differences, where multiple cases showed that appropriate location candidates could not be presented due to the distance between destinations and parking locations. The following improvements could address these challenges. First is the refinement of clustering methods. The current method faces issues such as the merging of densely packed stopping points in close proximity, or the combining of distant clusters causing cluster centers to deviate from intended locations. Introducing density-based clustering methods like OPTICS could enable more appropriate cluster formation [28]. Second is the utilization of spatial crowdsourcing, where tasks are assigned to participants who can physically visit locations.

This would enable identification of places that cannot be determined through map information alone [26].

B. Comparison between Crowdsourcing Results and Author Evaluation

To verify the consistency between CPoI evaluations obtained through crowdsourcing and the author's own assessments, we conducted a comparative analysis. The author examined each CPoI stopping position data individually to evaluate their suitability as cyclist stopping points. As shown in Figure 5, a positive correlation exists between the average crowdsourcing participant evaluations and the author's assessments, with many data points distributed along or near the diagonal line. This indicates general agreement between participant average ratings and author evaluations, confirming the basic effectiveness of the proposed evaluation method. However, several improvements could enhance evaluation accuracy. First, certain cases proved difficult to judge based solely on location names and Street View images. Features such as scenic quality and rest suitability require more comprehensive information presentation. Therefore, providing images from multiple perspectives and detailed information about surrounding environments should be considered. Additionally, addressing various heuristics that influence human judgment is important [29], [30]. Evaluations may be affected by anchoring effects where participants are influenced by initially presented criteria, availability heuristics based on personal experience, and representativeness heuristics leading to superficial judgments based on typical "tourist spot" characteristics. To address these challenges, establishing clearer evaluation criteria and introducing multifaceted assessments would be effective. Specifically, setting evaluation criteria from multiple concrete perspectives, such as "ease of bicycle parking," "quality of rest facilities," and "scenic appeal," could enable more objective evaluations.

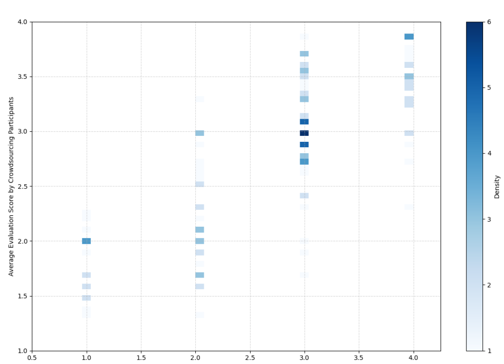

Fig. 5. Scatter Plot of Average Crowdsourcing Ratings versus Author Ratings

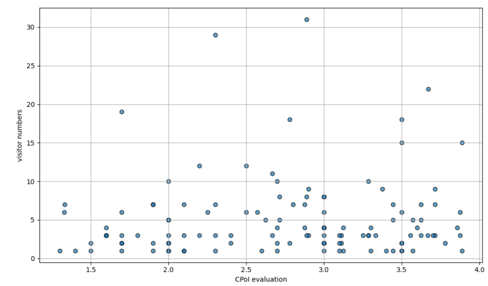

Fig. 6. Relationship between evaluation scores and number of visitors for each CPoI

C. Relationship Between CPoI Evaluation and Number of Visitors

Figure 6 shows the relationship between the number of GPS points in each CPoI cluster (number of visitors) and evaluation scores. This figure excludes the starting point where all cyclists stopped and five locations that were excluded from evaluation. The figure reveals that while low-rated CPoIs tend to have few visitors, highly-rated CPoIs show a wide distribution in visitor numbers. This distribution suggests the existence of Popular CPoIs and Hidden CPoIs. Although some locations with low ratings had many visitors, these were confirmed to be cases where location identification was unsuccessful due to discrepancies between destinations and parking locations. From the distribution of visitor numbers and evaluation scores, we can identify Popular CPoIs that attract sufficient numbers of cyclists and Hidden CPoIs that are highly rated despite relatively few visitors by using a threshold value α. In this study, we set $\alpha = 10$ and classified the CPoIs into Popular CPoIs with 10 or more visitors (shown in Table III) and Hidden CPoIs with fewer than 10 visitors (shown in Table IV). Popular CPoIs include tourist-oriented facilities such as the Dutch-style Windmill and Yoshioka Onsen Roadside Station, while Hidden CPoIs include locations where visitors can quietly enjoy local attractions, such as urban parks like Asahigaoka Park and historical buildings like Santai Shrine. It is generally known that tourist site visitor numbers correlate with distance to the destination, and the CPoIs identified in this study also showed a tendency for visitor numbers to decrease with increasing distance from the starting point. Considering this characteristic, more detailed analysis of CPoI features could be achieved through comparative analysis by distance zones or analysis that corrects for distance-related effects.

TABLE III
TOP 5 HIGHEST RATED POPULAR CPOIS

Rank	Place Name	visitor numbers	Rate
1	Dutch-style Windmill	15	3.89
2	Maebashi Akagi Roadside Station	30	3.67
2	Yoshioka Onsen Roadside Station Lot	22	3.67
4	Lunapark	18	3.50
4	Ushiya Kiyoshi	15	3.50

TABLE IV
TOP 5 HIGHEST RATED HIDDEN CPOIS

Rank	Place Name	visitor numbers	Rate
1	Asahigaoka Park	1	3.89
2	Santai Shrine	3	3.88
2	Chuo-Maebashi Station	6	3.88
4	Tokunaga Farm	4	3.86
5	Isshin-an	2	3.78

VI. CONCLUSION

In this study, we proposed a method for semi-automatically detecting and evaluating CPoIs from cyclists' ride log data. The proposed method identifies CPoIs through stopping point estimation and hierarchical clustering, and evaluates their characteristics using crowdsourcing. Through application to actual data, we identified 125 CPoIs and confirmed the effectiveness of crowdsourcing-based evaluation. Furthermore, through analyzing the relationship between evaluation scores and visitor numbers, we demonstrated the potential to distinguish between Popular CPoIs and Hidden CPoIs. These findings are expected to contribute to the discovery of new tourism resources in cycle tourism. Future challenges include improving clustering methods, clarifying evaluation criteria, and analyzing the effects of distance. We expect that the outcomes of this research will contribute to the development of sustainable tourism through cycle tourism.

REFERENCES

[1] B. W. Ritchie, "Bicycle tourism in the south island of new zealand: Planning and management issues," *Tourism management*, vol. 19, no. 6, pp. 567–582, 1998.

[2] C. G. Prato, K. Halldórsdóttir, and O. A. Nielsen, "Evaluation of land-use and transport network effects on cyclists' route choices in the copenhagen region in value-of-distance space," *International journal of sustainable transportation*, vol. 12, no. 10, pp. 770–781, 2018.

[3] R. Khatri, C. R. Cherry, S. S. Nambisan, and L. D. Han, "Modeling route choice of utilitarian bikeshare users with gps data," *Transportation research record*, vol. 2587, no. 1, pp. 141–149, 2016.

[4] D. M. Scott, W. Lu, and M. J. Brown, "Route choice of bike share users: Leveraging gps data to derive choice sets," *Journal of Transport Geography*, vol. 90, p. 102903, 2021.

[5] P. Chen, Q. Shen, and S. Childress, "A gps data-based analysis of built environment influences on bicyclist route preferences," *International journal of sustainable transportation*, vol. 12, no. 3, pp. 218–231, 2018.

[6] M. Zimmermann, T. Mai, and E. Frejinger, "Bike route choice modeling using gps data without choice sets of paths," *Transportation Research Part C-emerging Technologies*, vol. 75, pp. 183–196, 2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:55096451

[7] B. Thomee and A. Rae, "Uncovering locally characterizing regions within geotagged data," in *Proceedings of the 22nd international conference on World Wide Web*, 2013, pp. 1285–1296.

[8] C.-L. Kuo, T.-C. Chan, I.-C. Fan, and A. Zipf, "Efficient method for poi/roi discovery using flickr geotagged photos," *ISPRS International Journal of Geo-Information*, vol. 7, no. 3, p. 121, 2018.

[9] Y. Kozaki, Y. Wang, and Y. Kawai, "Generating pictorial maps for tourists using flickr photo data," in *2018 IEEE 7th Global Conference on Consumer Electronics (GCCE)*. IEEE, 2018, pp. 403–407.

[10] C. M. Krause and L. Zhang, "Short-term travel behavior prediction with gps, land use, and point of interest data," *Transportation Research Part B: Methodological*, vol. 123, pp. 349–361, 2019.

[11] D. Ashbrook and T. Starner, "Using gps to learn significant locations and predict movement across multiple users," *Personal and Ubiquitous computing*, vol. 7, pp. 275–286, 2003.

[12] Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma, "Mining interesting locations and travel sequences from gps trajectories," in *Proceedings of the 18th international conference on World wide web*, 2009, pp. 791–800.

[13] D. Kitayama, "Extraction method for anaba spots based on name recognition and user's evaluation," in *Proceedings of the 18th International Conference on Information Integration and Web-based Applications and Services*, 2016, pp. 12–15.

[14] C. Zhuang, Q. Ma, X. Liang, and M. Yoshikawa, "Discovering obscure sightseeing spots by analysis of geo-tagged social images," in *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, 2015, pp. 590–595.

[15] ——, "Anaba: An obscure sightseeing spots discovering system," in *2014 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2014, pp. 1–6.

[16] C. Zhuang, Q. Ma, and M. Yoshikawa, "Sns user classification and its application to obscure poi discovery," *Multimedia Tools and Applications*, vol. 76, pp. 5461–5487, 2017.

[17] Y. Yang, Y. Duan, X. Wang, Z. Huang, N. Xie, and H. T. Shen, "Hierarchical multi-clue modelling for poi popularity prediction with heterogeneous tourist information," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 4, pp. 757–768, 2018.

[18] Y. Cui, L. Deng, Y. Zhao, B. Yao, V. W. Zheng, and K. Zheng, "Hidden poi ranking with spatial crowdsourcing," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 814–824.

[19] X. Liu and Y. Long, "Automated identification and characterization of parcels with openstreetmap and points of interest," *Environment and Planning B: Planning and Design*, vol. 43, no. 2, pp. 341–360, 2016.

[20] L. Huang, Q. Li, and Y. Yue, "Activity identification from gps trajectories using spatial temporal pois' attractiveness," in *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on location based social networks*, 2010, pp. 27–30.

[21] D. Zhao, W. Zhou, W. Wang, and X. Hua, "Trip purpose prediction using travel survey data with poi information via gradient boosting decision trees," *IET Intelligent Transport Systems*, vol. 18, no. 2, pp. 269–289, 2024.

[22] S. Lyu, T. Han, P. Li, X. Luo, and T. Kusakabe, "A dual-flow attentive network with feature crossing for chained trip purpose inference," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 1, pp. 631–644, 2022.

[23] T. Feng and H. H. Timmermans, "Detecting activity type from gps traces using spatial and temporal information," *European Journal of Transport and Infrastructure Research*, 2015. [Online]. Available: https://api.semanticscholar.org/CorpusID:5392726

[24] D. Dhananjaya and T. Sivakumar, "Enhancing the poi data for trip purpose inference using machine learning techniques," in *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE Press, 2022, p. 3496–3501. [Online]. Available: https://doi.org/10.1109/ITSC55140.2022.9922534

[25] J. H. Ward Jr, "Hierarchical grouping to optimize an objective function," *Journal of the American statistical association*, vol. 58, no. 301, pp. 236–244, 1963.

[26] Z. Chen, R. Fu, Z. Zhao, Z. Liu, L. Xia, L. Chen, P. Cheng, C. C. Cao, Y. Tong, and C. J. Zhang, "gmission: A general spatial crowdsourcing platform," *Proceedings of the VLDB Endowment*, vol. 7, no. 13, pp. 1629–1632, 2014.

[27] G. Foody, L. See, S. Fritz, M. Velde, C. Perger, C. Schill, and D. Boyd, "Assessing the accuracy of volunteered geographic information arising from multiple contributors to an internet based collaborative project," *Transactions in GIS*, vol. 17, 2013.

[28] I. Semanjski, M. Ramachi, and S. Gautama, "Detection of points of interest from crowdsourced tourism data," in *Computational Science and Its Applications–ICCSA 2019: 19th International Conference, Saint Petersburg, Russia, July 1–4, 2019, Proceedings, Part II 19*. Springer, 2019, pp. 203–216.

[29] K. Burghardt, T. Hogg, R. D'Souza, K. Lerman, and M. Pósfai, "Origins of algorithmic instabilities in crowdsourced ranking," *Proceedings of the ACM on Human-Computer Interaction*, vol. 4, pp. 1 – 20, 2020.

[30] K. Burghardt, T. Hogg, and K. Lerman, "Quantifying the impact of cognitive biases in question-answering systems," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 12, no. 1, 2018.