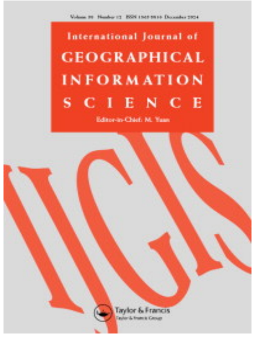


Taylor & Francis
Taylor & Francis Group

International Journal of Geographical Information Science

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tgis20

Informed sampling and recommendation of cycling routes: leveraging crowd-sourced trajectories with weighted-latent Dirichlet allocation

Weilian Li, Jan-Henrik Haunert, Axel Forsch, Jun Zhu, Qing Zhu & Youness Dehbi

To cite this article: Weilian Li, Jan-Henrik Haunert, Axel Forsch, Jun Zhu, Qing Zhu & Youness Dehbi (2024) Informed sampling and recommendation of cycling routes: leveraging crowd-sourced trajectories with weighted-latent Dirichlet allocation, International Journal of Geographical Information Science, 38:12, 2492-2513, DOI: 10.1080/13658816.2024.2391428

To link to this article: https://doi.org/10.1080/13658816.2024.2391428

Published online: 23 Sep 2024.

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
2024, VOL. 38, NO. 12, 2492–2513
https://doi.org/10.1080/13658816.2024.2391428

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Informed sampling and recommendation of cycling routes: leveraging crowd-sourced trajectories with weighted-latent Dirichlet allocation

Weilian Li[a,b], Jan-Henrik Haunert[c], Axel Forsch[c], Jun Zhu[b], Qing Zhu[b] and Youness Dehbi[a]

[a]Computational Methods Lab, HafenCity University Hamburg, Hamburg, Germany; [b]Faculty of Geosciences and Engineering, Southwest Jiaotong University, Chengdu, China; [c]Institute of Geodesy and Geoinformation, University of Bonn, Bonn, Germany

ABSTRACT

Attractive cycling routes can effectively promote active mobility, thus reducing the twin pressures of the population boom and the greenhouse effect. However, the existing approaches for cycling route recommendation primarily concentrate on identifying the most efficient routes while ignoring the urban spatial context, which is essential to meet the user's particular preferences. This article proposes a novel method for informed sampling and recommending cycling routes leveraging crowd-sourced trajectories with weighted-latent Dirichlet allocation (WLDA). Precisely, spatial context mapping, incorporating a weighting mechanism into LDA, latent topics mining, and cycling route recommendation based on informed sampling are introduced. We collected 1,016 cycling trajectories around Cologne, Germany, for experimental analysis. The experimental results show that the three latent topics within the trajectories, leisure, city, and green tours, are clearly presented in the line density analysis. The insightful recommendation for unfamiliar cyclists could also be actively sampled upon the WLDA model. These findings suggest that our approach could shift the route recommendation paradigm from GIS analysis to a semantic mining perspective, yielding highly interpretable results and offering novel research avenues for applying machine learning in route planning.

ARTICLE HISTORY
Received 21 April 2024
Accepted 7 August 2024

KEYWORDS
Cycling route recommendation; weighted-latent Dirichlet allocation; crowd-sourced trajectories; spatial context mapping; natural language processing

1. Introduction

Urban transportation is gradually transforming toward sustainability, driven by the twin pressures of the population boom and the greenhouse effect, and green travel is becoming people's preferred choice (Chen *et al.* 2018, Yang *et al.* 2021, 2020, Meng and Zheng 2023). Particularly, cycling is receiving increasing attention as an alternative to public transport, which was more evident during the pandemic since people

CONTACT Youness Dehbi youness.dehbi@hcu-hamburg.de

attempted to avoid crowds and increase contactless outdoor activities (Buehler and Pucher 2021, Dingil and Esztergár-Kiss 2021, Guo *et al.* 2021, Hu *et al.* 2021).

Significantly, most governments support expanding and improving cycling infrastructure, programs, and policies to ensure cycling thrives (Buehler and Pucher 2021). In general, cycling has the following benefits: (a) the protection of the environment by diminishing noise, air pollution, and carbon emissions (Eren and Uz 2020, Ding *et al.* 2021), (b) the reduction of congestion in the transport network and parking facilities (Pucher and Buehler 2017), and (c) the promotion of public health through physical activity (Ogilvie *et al.* 2011, Teixeira *et al.* 2021).

Attractive cycling routes can effectively promote active mobility (Huang *et al.* 2015, Martín and Páez 2019, Meng and Zheng 2023). For example, Fraser and Lock (2011) clarified that there are statistically significant associations between green and recreational spaces and cycling. Harms *et al*. (2014) stated that the attractiveness of the built environment along cycling routes positively affects cycling levels. Similarly, Biassoni *et al*. (2023) revealed that the perception of road infrastructures and environmental attitudes have a positive impact on cycling frequency. Therefore, a suitable route can play a crucial role in encouraging more people to adopt cycling as a mode of transportation.

In recent years, recreational cycling has gradually gained attention compared with utility and commuting cycling, particularly since the pandemic has created new momentum (Abdullah *et al.* 2021, Buehler and Pucher 2021). Recreational cycling refers to cycling for leisure, health, or fitness, focusing more heavily on attractive natural or artificial environments (Etminani-Ghasrodashti *et al.* 2018). People do not care much about the origin or destination, and they may even take the train or car to the recreational areas and start cycling there rather than riding a bicycle for the entire trip (Nguyen and Pojani 2022). In this context, the most well-known navigation app, Google Maps, is not very helpful for cyclists to choose a suitable route since it is programmed to find, e.g. the most cost-effective route instead of the most beautiful one (Siriaraya *et al.* 2020). In addition to Google Maps, there are several popular bicycle routing and advocacy tools, such as Open Source Routing Machine (OSRM)[1] and CycleStreets.[2] OSRM is an open-source router designed for shortest path planning in road networks. CycleStreets is a free cycle journey planner that can currently infer the routes that are rated fastest, quietest, or balanced (Desjardins *et al.* 2022, Lovelace *et al.* 2022). According to the website, the CycleStreets platform is able to plan leisure tours, this is however restricted to admin users right now.

In the context of Volunteered Geographic Information (VGI), large sets of crowdsourced trajectories are collected, which cover various riding scenarios and implicitly reflect cyclists' behaviors and preferences compared to newly planned cycling routes (Goodchild 2007, Sultan *et al.* 2017, Oehrlein *et al.* 2018, Feng *et al.* 2020, Yan *et al.* 2020, Fang *et al.* 2022, Forsch *et al.* 2023). Nowadays, some cycling apps (e.g. Komoot, Bikemap) can suggest routes matching the user's preferences (e.g. difficulty, preferred surface) by selecting a subset from a large dataset of trajectories rather than planning a new route. However, they cannot tell the cyclist that the recommended route is closer to nature or more recreational, which is necessary to meet the user's particular needs.

From the authors' point of view, the VGI-based cycling route recommendation for recreational cycling cannot mainly look at the trajectory itself while ignoring the spatial context and underlying semantics of road segments. Trajectories weave through a network of streets, embracing both natural landscapes and artificial structures. Aggregating this geospatial information into a corpus is able to semantically reveal cyclists' preferences and underlying cycling themes (Chu *et al.* 2014, Gao *et al.* 2023). In this context, we aim to innovatively transform the route recommendation from a GIS analysis to a semantic mining perspective. Specifically, this research does not aim to plan a new cycling route but seeks to retrieve a subset of routes from a large dataset of trajectories matching the latent users' preferences and recommend a route with a specific topic for cyclists building upon.

The study workflow of this article is shown in Figure 1. Based on a record of VGI-collected cyclist trajectories, a map matching with the underlying street network has been performed (Behr *et al.* 2021). This allows for extracting further attributes reflecting semantic knowledge, such as the road attributes in the neighborhood. Such knowledge is translated into a vector-based description of the share of each attribute. This paves the way for using Latent Dirichlet allocation (LDA) to induce the latent trajectory topics by learning an underlying probability distribution. An innovative expansion of LDA by a weighting mechanism, which we call WLDA, allows for an informed sampling and recommendation of routes depending on the a-priori known cyclist type from the underlying parametric distribution.

The remainder of this article is structured as follows: Section 2 reviews the related work of this study. Section 3 introduces the experimental data. Section 4 provides a detailed explanation of the proposed method. Section 5 analyzes the experimental results. Section 6 presents the discussion. Section 7 concludes the main contributions and gives an outlook for future research.

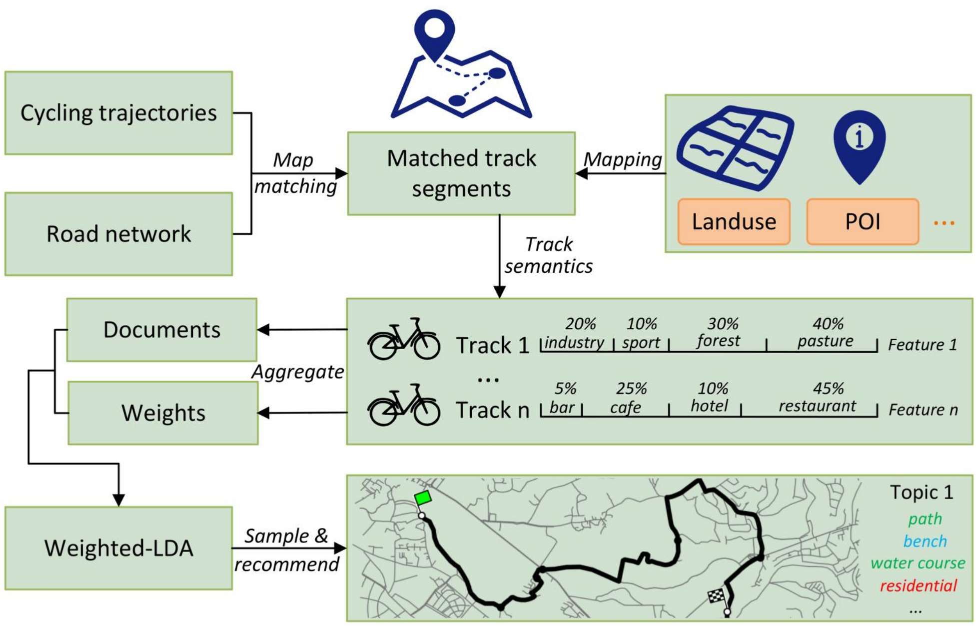


Figure 1. Overview of the work process. We start by mapping the spatial context into road segments using spatial statistical analysis. Subsequently, each trajectory is regarded as a meaningful textual document comprising words and associated weights. Finally, WLDA is used to extract hidden topics of cycling semantics, which could help cyclists choose the desirable cycling routes according to their preferences.

2. Related work

2.1. Route recommendation

Route recommendation usually suggests planning a route between two specific points in a road network (de Oliveira e Silva *et al.* 2022). Early studies on route recommendation focused on discovering the most efficient routes according to a particular metric (Cui *et al.* 2018, Meng and Zheng 2023), such as the shortest distance. This approach is programmed to find the shortest route while ignoring the need for active mobility and user preferences. Subsequently, multiple criteria GIS analysis is becoming popular by considering different indicators (e.g. time, distance, slope) to plan routes that meet cyclists' benefits (Đerek and Sikora 2019, Li *et al.* 2021, 2024a, Zhu 2022). The most striking feature of multiple criteria GIS analysis is that experts generally define the weight of indicators through a pairwise comparison in which preference factors might not reflect the cyclist's actual behaviors (Nadi and Delavar 2011, de Oliveira e Silva *et al.* 2022, Nasution *et al.* 2022, Li *et al.* 2024b). Likewise, both OSRM and CycleStreets support user-defined profiles to plan a new cycling route by setting different metrics weights (Mahfouz *et al.* 2023).

In terms of route recommendation for recreational cycling, people are motivated mainly by a desire to enjoy scenic views or physical exercise (Chen and Chen 2013), and they may not care much about the origin or destination but rather a cycling route that meets their preferences. In addition, cyclists often utilize smartphones or smartwatches with built-in GNSS sensors to track their activities, and this results in sizeable trajectory data (Brauer *et al.* 2021). Trajectory originally is defined by the Cambridge Dictionary[3] as 'the curved path that an object follows after it has been thrown or shot into the air', but in our case, it refers to a set of sampled GPS positions with a time stamp.

The trajectories generated by users exhibit more comprehensive spatiotemporal coverage that implicitly reflects cyclists' behaviors and preferences, which can be learned and modeled from the historical GPS trajectories. For instance, Cui *et al.* (2018) used matrix factorization (MF) to estimate the users' travel behavior probabilities based on the travel frequencies obtained from the trajectories. Subsequently, the route with maximum travel probability was generated using the naïve Bayes model. Similarly, Mou *et al.* (2022) utilized geo-tagged Flickr photos to generate user travel trajectories, which were further inputted into a recurrent neural network for personalized travel route recommendations. Nevertheless, these studies considered the user's behavior (e.g. travel time and frequency) but did not fully incorporate the spatial context, and the interpretability of their results remains limited.

2.2. Latent Dirichlet allocation

Latent Dirichlet allocation (LDA) is a generative probabilistic model that can infer topics from previously unseen documents (Blei *et al.* 2003). Unlike k-means or principal component analysis (PCA), LDA allows a word to belong to several clusters with varying degrees rather than inducing only distinct clusters (Imran *et al.* 2015, Wang and Ye 2018) and is considered the state-of-art approach for topic modeling (Nugroho *et al.* 2020).

Topic modeling effectively extracts semantic topics or patterns from texts (Liu *et al.* 2019, Li *et al.* 2023). Since 2019, many researchers have mined and analyzed public opinion, vaccination intentions, and rumors related to COVID-19 based on LDA (Han *et al.* 2020, AlAgha 2021). However, the existing studies focus on simply applying LDA to deal with social media comments but rarely on spatial data mining, particularly GPS trajectories. To our knowledge, some researchers adopted LDA to extract topics from massive taxi trajectories by considering each trajectory as a document, which facilitates exploring various movement patterns of users (Chu *et al.* 2014, Tang *et al.* 2018, Liu *et al.* 2019, Gao *et al.* 2023). The above studies mapped GPS coordinates to street names and then directly fed them into LDA to extract hidden topics without considering the influence of the spatial context on cyclist behavior. In addition, word frequency is predominant in the current LDA modeling process rather than word importance. In contrast, this article incorporates a weighting mechanism into LDA, which allows for sampling and recommending cycling routes from a large dataset of crowd-sourced trajectories.

3. Data

Figure 2 shows the data used in this study. The trajectory data was provided by Forsch *et al.* (2023), which originated from a user-driven platform known as GPSies.[4] This platform has diverse groups of cyclists and features a wide range of cycling activities, including scenic bike touring, mountain biking, and road biking. Consequently, we have grounds to assume that these cyclists have distinct preferred intentions for their cycling pursuits. The 1,016 trajectories were recorded by cyclists around Cologne, Germany, in a search radius of 25 km. Each trajectory is associated with a track ID but no other user information.

Meng and Zheng (2023) stated that an attractive cycling route should consider accessibility, suitability, and visual perceptibility. Therefore, we collected POI, road types, and landuse data accordingly. The POI and road types data were derived from the OpenStreetMap (OSM),[5] and landuse data were obtained from Open North-Rhine-Westphalia (NRW).[6]

4. Methodology

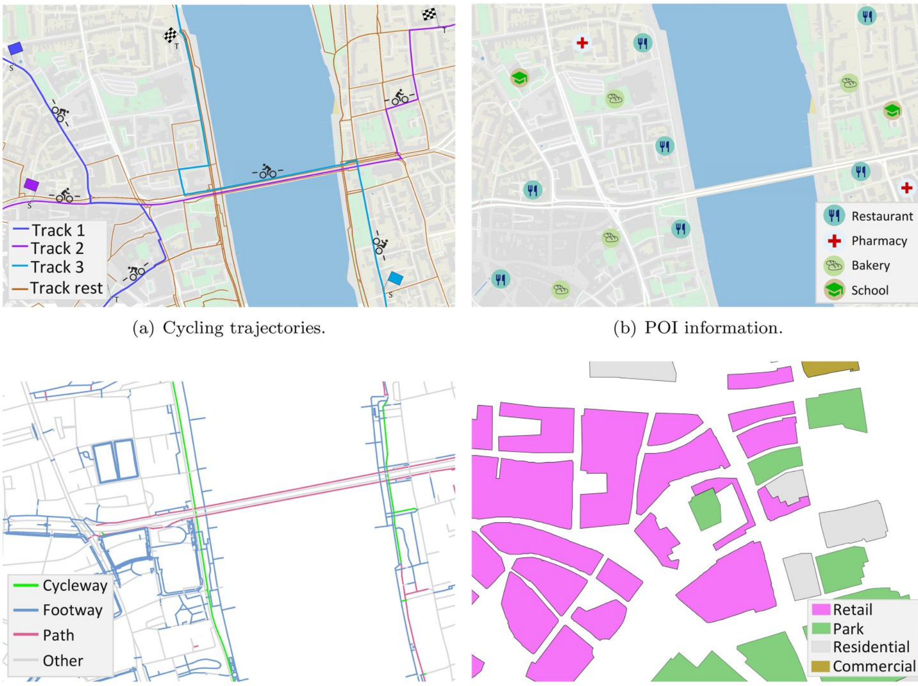
This section describes the methodology we use for informed sampling and recommendation of cycling routes. We start by mapping spatial context into the trajectory, then introduce WLDA for topic mining of trajectories and evaluation, and finally illustrate the cycling routes recommendation based on topic probability.

4.1. Mapping spatial context into trajectory

4.1.1. POI integration

Figure 3 shows the integration process of POI and trajectory data. Initially, we create a buffer with a radius of 50 m around the trajectory, which is used to intersect with the POI layer. Within this buffer, the number of POI is counted by category. Equation (1) defines w_{ct} as the proportion of POI category c associated with a trajectory t.

$$w_{ct} = \frac{|P_c \cap B_t|}{|P \cap B_t|}, P_c \in P \tag{1}$$



(a) Cycling trajectories. (b) POI information. (c) Road type. (d) Landuse category.

Figure 2. The primary data of this work. (a) Cycling trajectories, the total length of trajectories up to 34,739 km. (b) POI information, including restaurant, bakery, school, etc. (c) Road type data, including cycleway, footway, path, etc. (d) Landuse data, including park, retail, residential, etc.

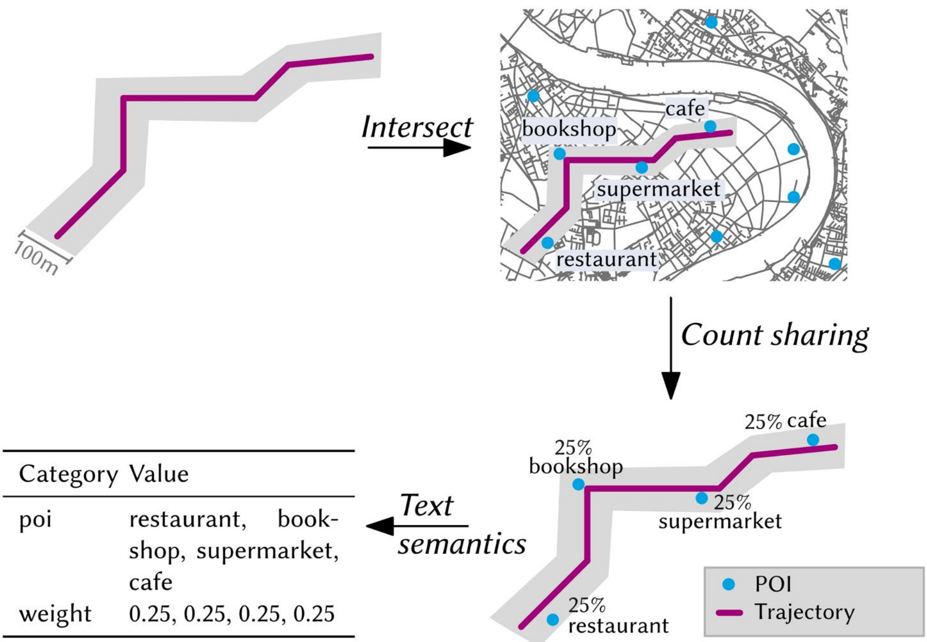


Figure 3. The integration process of POI and trajectory data.

where P indicates the point set of all POI categories, P_c represents a point set with a given category c, and B_t denotes the buffer trajectory t. Finally, the POI categories and their corresponding shares are used as one of the spatial features of the trajectory.

4.1.2. Road semantics mapping

The cycling trajectories after map matching lack semantic information, but their spatial location aligns with the road network. Our objective is to join the road types to the trajectory by location and calculate the length of road segments, as illustrated in Figure 4.

In addition, Equation (2) calculates the proportion w_{st} of a road type s in a given trajectory t.

$$w_{st} = \frac{\mathcal{L}(r_s)}{\mathcal{L}(t)}, \tag{2}$$

where $\mathcal{L}$ is used to calculate the length of a line geometry. r_s represents a road line with type s, and t is a given trajectory. The road types and their proportions can further enrich the trajectory semantics. Mapping road semantics to trajectories can characterize cycling suitability.

4.1.3. Landuse sharing

The landuse class can implicitly reflect the visual perceptual preferences of cyclists. Likewise, we follow the workflow in Figure 5 to compute the landuse sharing with the

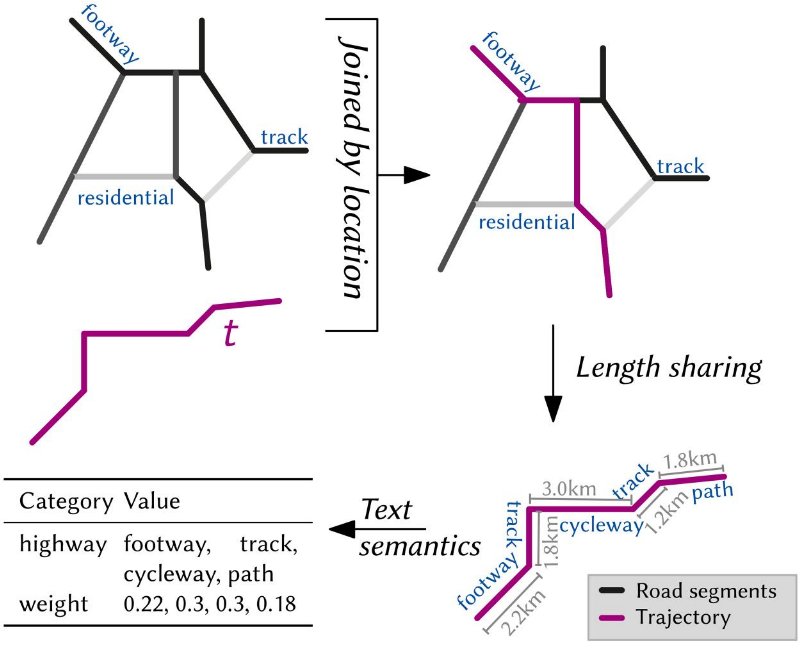


Figure 4. Mapping road semantics to the trajectory.

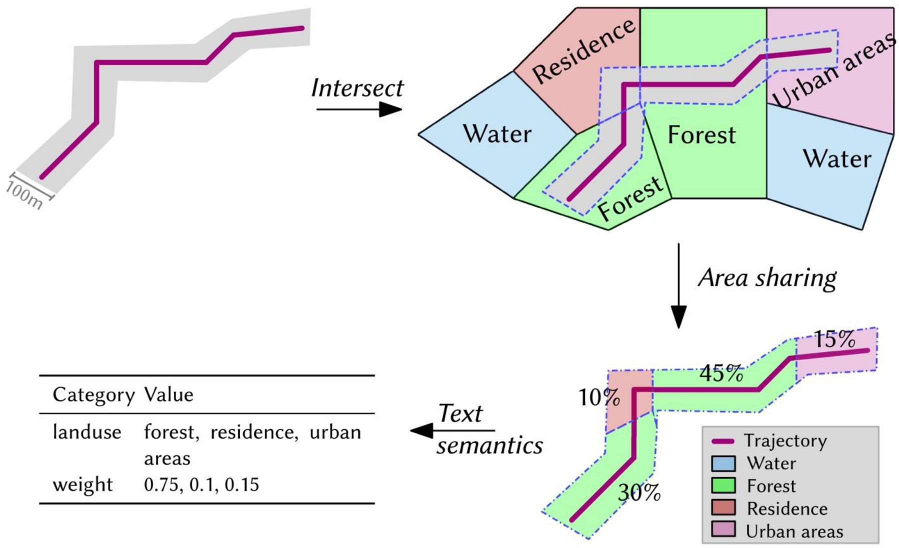


Figure 5. Landuse sharing with the trajectory.

trajectory buffer. Specifically, Equation (3) is used to determine the sharing w_{lt} of each landuse class l within in the buffer of a given trajectory t.

$$w_{lt} = \frac{\mathcal{A}(A_l \cap B_t)}{\mathcal{A}(B_t)}, \tag{3}$$

where $\mathcal{A}$ is used to calculate the area of polygonal geometries, A_l is the surface coverted by landuse class l, and B_t refers to the buffer trajectory t. Furthermore, the landuse classes and their associated contribution are translated into text semantics.

4.1.4. Weight normalization

This work employs a weighted linear combination to portray each spatial feature's contribution in specifying the cycling trajectory's attractiveness, as shown in Equation (4).

$$w = \gamma_1 \cdot w_{ct} + \gamma_2 \cdot w_{st} + \gamma_3 \cdot w_{lt} \tag{4}$$

In this equation, w indicates the weighted contribution of spatial context, γ_1, γ_2, and γ_3 is the weight of w_{ct}, w_{st}, and w_{lt}, respectively. The sum of γ_1, γ_2, and γ_3 equals 1. In our case, w_{ct}, w_{st}, and w_{lt} are set to 0.3, 0.3, and 0.4.

4.2. Latent topics mining in cycling trajectories

In this section, we introduce how to model WLDA and use it to mine the latent topics in the cycling trajectories. Specifically, WLDA for topic modeling of trajectories, word-weighted Gibbs sampling, and topic evaluation based on Jensen-Shannon divergence are described in detail.

4.2.1. WLDA for topic modeling of trajectories

4.2.1.1. Principles of WLDA modeling. Figure 6 shows the principles of WLDA modeling and the workflow for trajectory topic mining. In this diagram, K denotes the number of topics. α and β are the Dirichlet prior for topic and word distribution, respectively. The process of generating topics from the WLDA model is as follows:

- Trajectory semanticization. The spatial context (e.g. landuse, road type) associated with trajectories is semanticized to keywords and weights.
- Trajectory textualization. Trajectories and their spatial context are converted into a textual corpus. Here, we regard each trajectory as a meaningful document comprising words of spatial context and associated weights.
- Topic modeling. For the document T_m, the topic distribution is $\theta_m \sim \mathrm{Dir}(\alpha)$. For the topic Z_k, the word distribution is $\varphi_k \sim \mathrm{Dir}(\beta)$. First, randomly generate a topic $z_{mn} \sim \mathrm{Multinomial}(\theta_m)$. Second, randomly extract a word $w_{mn} \sim \mathrm{Multinomial}(\varphi_{z_{mn}})$.
- Random sampling. Gibbs sampling, a Markov chain Monte Carlo (MCMC) algorithm for sampling from a specified multivariate probability distribution (Gelfand 2000), is used to infer the probability p of a word being assigned a topic and estimate implicit parameters θ_m and φ_k. The detailed process will be illustrated in Section 4.2.2.
- Result output. Output track-topic and topic-word probability as a basis for the subsequent trajectory recommendation.

4.2.1.2. Algorithm outline. As input, the WLDA algorithm requires a textual corpus $\mathcal{CP}$, which comprises a set of trajectories T with spatial context and a collection of the features' weights W. Formally, $\mathcal{CP} = \{(T_m^n, W_m^n) | 1 \le m \le M, 1 \le n \le N\}$, M corresponds to the number of underlying trajectories, whereas N is the number of words in a given trajectory, which could vary in different trajectories. T_m^n indicates the n^{th} word in the m^{th} trajectory, W_m^n denotes the corresponding weight of the word T_m^n. The algorithm's output includes sets θ and φ. For $\theta_m \in \theta$ represents the track-topic distribution of the m^{th} trajectory. Similarly, $\varphi_k \in \varphi$ indicates the topic-word distribution of the k^{th} topic. The generic flow of WLDA is summarized in Algorithm 1.

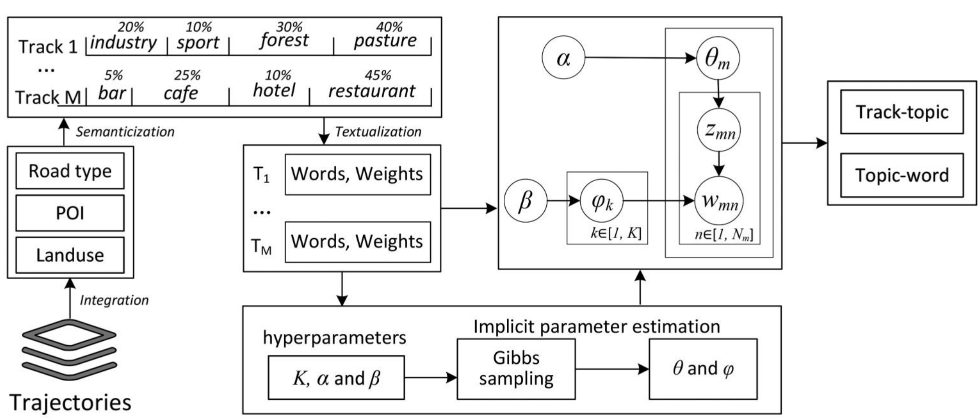


Figure 6. Diagram of WLDA for topic modeling of trajectories.

Algorithm 1. The generic flow of WLDA.

Input: trajectory corpus $\mathcal{CP} = \{(T_m^n, W_m^n) | 1 \le m \le M, 1 \le n \le N\}$, topic number K, m^{th} track-topic weights matrices δ_m^k and cumulative weights matrices δ_m, k^{th} topic-word weights matrices η_k^v and cumulative weights matrices η_k, number of iterations X

Output: estimated m^{th} track-topic distribution θ_m and k^{th} topic-word distribution φ_k

for $1 \le m \le M$ **do** // Randomly initialize the topic
  for each *word* T_m^n **do**
    $k \leftarrow$ random(0, K);
    $T_m^n \leftarrow k$;
    $\delta_m^k \leftarrow \delta_m^k + W_m^n$;
    $\delta_m \leftarrow \delta_m + W_m^n$ // Increment weights for the new assigned topic;
    $\eta_k^v \leftarrow \eta_k^v + W_m^n$;
    $\eta_k \leftarrow \eta_k + W_m^n$;
while $x < X$ **do** // Sample new topics
  for $1 \le m \le M$ **do**
    for each *word* T_m^n **do**
      $T_m^n \leftarrow$ Sampling
estimate θ_m and φ_k // Return track-topic and topic-word distributions

4.2.2. Word-weighted Gibbs sampling

4.2.2.1. Sampling process. The parameter estimation of WLDA is a complex optimization problem. In our case, we propose word-weighted Gibbs sampling to estimatec-process of which is summarized in Algorithm 2.

Algorithm 2. Gibbs sampling for WLDA.

Input: topic k, track-topic weights matrix δ_m^k, topic-word weights matrix η_k^v, topic numbers K

Output: updated k, δ_m^k, and η_k^v

$k \leftarrow T_m^n$ // Assign the initial topic;
$\delta_m^k \leftarrow \delta_m^k - W_m^n$;
$\delta_m \leftarrow \delta_m - W_m^n$ // Less weights for the new assigned topic;
$\eta_k^v \leftarrow \eta_k^v - W_m^n$;
$\eta_k \leftarrow \eta_k - W_m^n$;
calculate the probability p;
for $k < K$ **do** // Update topic
  if $p[k] > random(0, \max(p))$ **then**
    break;
$\delta_m^k \leftarrow \delta_m^k + W_m^n$;
$\delta_m \leftarrow \delta_m + W_m^n$ // Increment weights for the new assigned topic;
$\eta_k^v \leftarrow \eta_k^v + W_m^n$;
$\eta_k \leftarrow \eta_k + W_m^n$;
return k, δ_m^k, and η_k^v

The conditional probability p of a word being assigned a topic can be calculated using Equation (5).

$$p = \frac{\delta_m^k + \alpha_k}{\sum_{k=1}^{K}\left(\delta_m^k + \alpha_k\right)} \cdot \frac{\eta_k^v + \beta_v}{\sum_{v=1}^{V}\left(\eta_k^v + \beta_v\right)} \tag{5}$$

where, δ_m^k represents the cumulative weights of the m^{th} trajectory being assigned topic k. η_k^v denotes the cumulative weights of the v^{th} word in the k^{th} topic. K is the number of topics, V is the number of words being assigned topics. α_k and β_v are hyperparameters.

4.2.2.2. Parameter estimation. Furthermore, Equation (6) can be employed to estimate the model parameters θ and φ.

$$\begin{cases} \theta_m^k = \dfrac{\delta_m^k + \alpha_k}{\sum_{k=1}^{K}\left(\delta_m^k + \alpha_k\right)} \\ \varphi_k^v = \dfrac{\eta_k^v + \beta_v}{\sum_{v=1}^{V}\left(\eta_k^v + \beta_v\right)} \end{cases} \tag{6}$$

where θ_m^k indicates the m^{th} track with a k^{th} topic probability, φ_k^v indicates the k^{th} topic with a v^{th} word probability. We can get the topic distributions of trajectories, $\theta = [\theta_1, \ldots, \theta_m, \ldots, \theta_M | 1 \le m \le M]$, $\theta_m = \left[\theta_m^1, \ldots, \theta_m^k, \ldots, \theta_m^K | 1 \le k \le K\right]$, e.g. $\theta_1 = \{$topic1: 80%, topic2: 10%, $\ldots\}$. Likewise, the word distributions of topics, $\varphi = [\varphi_1, \ldots, \varphi_k, \ldots, \varphi_K | 1 \le k \le K]$, $\varphi_k = \left[\varphi_k^1, \ldots, \varphi_k^v, \ldots, \varphi_k^V | 1 \le v \le V\right]$, e.g. $\varphi_1 = \{$cycleway: 10%, farmland: 15%, $\ldots\}$.

4.2.3. Topic evaluation

Given that WLDA is an unsupervised method, the number of topics significantly influences the modeling outcomes. Determining an optimal number of topics relies on the perplexity measure (Zhang *et al.* 2021, Li *et al.* 2023). However, this metric tends to overestimate the parameter in the case of high sparsity, leading to unreliable estimates (Neishabouri and Desmarais 2020). In this context, we use the Jensen-Shannon divergence (JSD) to measure the semantic similarity between different distributions of topics, as shown in Equation (7). A smaller JSD value indicates a more significant similarity between two probability distributions $P\&Q$.

$$JSD(P||Q) = \frac{1}{2}D(P||M) + \frac{1}{2}D(Q||M), \tag{7}$$

where M is a mixture distribution equal to $\frac{1}{2}(P+Q)$. D measures the Kullback–Leibler divergence between two probability distributions.

4.3. Cycling routes recommendation based on topic probability

Figure 7 illustrates a workflow for generating a route recommendation building upon topic probability, allowing the selection of a subset of routes matching the users'

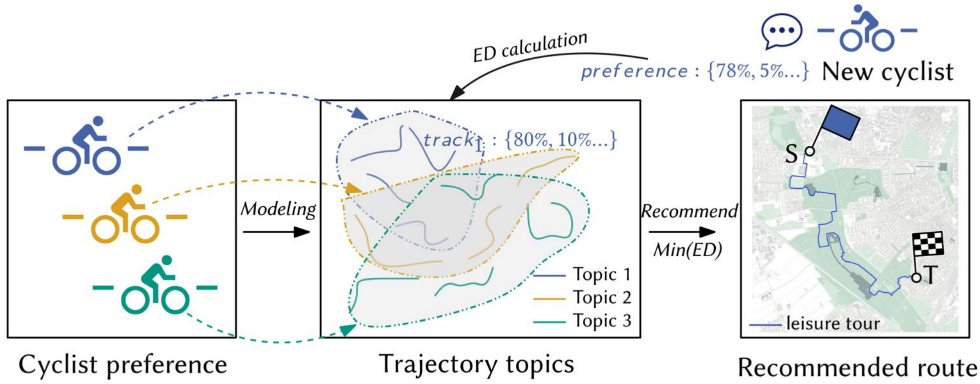


Figure 7. A workflow for generating a route recommendation building upon topic probability.

preferences. The latter represents prior knowledge for triggering an informed sampling of adequate trajectories.

Suppose the cyclist's preference vector $u = [u_1, \ldots, u_k, \ldots, u_K | 1 \le k \le K]$, K is the number of topics. Topic distributions of trajectories $\theta = [\theta_1, \ldots, \theta_m, \ldots, \theta_M | 1 \le m \le M]$, $\theta_m = \left[\theta_m^1, \ldots, \theta_m^k, \ldots, \theta_m^K | 1 \le k \le K\right]$, M corresponds to the number of underlying trajectories. The minimal Euclidean distance (ED) $d_{\min}$ between u and θ can be calculated using Equation (8).

$$d_{\min} = argmin\sqrt{\sum_{k=1}^{K}\left(\theta_m^k - u_k\right)^2} \tag{8}$$

where θ_m^k indicates the m^{th} track with a k^{th} topic probability, u_k indicates the user's preference for the k^{th} topic. In case of equal minimal distances, we select the topic with the highest percentage of user-specified preferences.

5. Results

In the following, we present our experiments and discuss their outcomes. First, we test the optimal number of topics set and analyze the topic clusters of trajectories. Subsequently, we perform line density analysis on trajectories to visualize the spatial distribution of cycling topics. Finally, we achieve the cycling route recommendation based on topic probability.

5.1. Topic analysis

5.1.1. The number of topics set

To determine the optimal number of topics K, we calculated the maximum, minimum, and average values of *JSD* from 2 to 10 topics, as shown in Figure 8.

We observed that when the number of topics exceeds 3, the JSD minimum distance between topics undergoes a steep drop, which indicates that at least two topics exhibit high similarity. In contrast, when $K = 3$, the JSD maximum, minimum, and

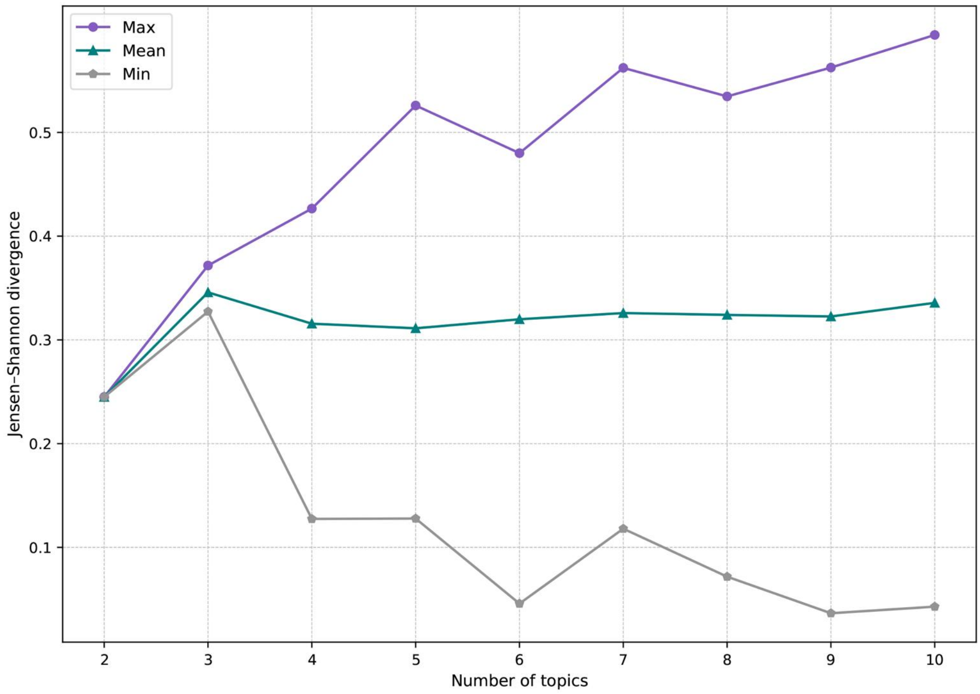


Figure 8. JSD distance between topics.

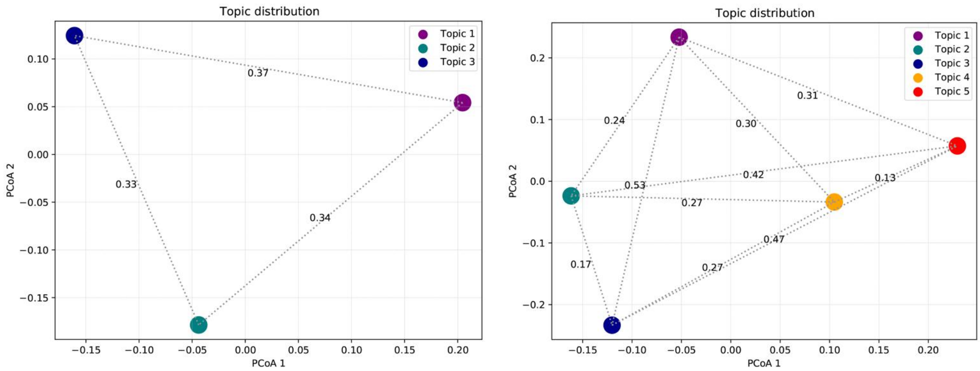
average distances are close to each other. This implies that the distribution of topics is well-dispersed and has a better classification result.

In addition, we took $K = 3$ and $K = 5$ as examples and used Principal Coordinate Analysis (PCoA) to visually portray the similarity between topics, as shown in Figure 9. Notably, when $K = 5$, a pronounced similarity is observed between Topic 4 and Topic 5. Considering the above results and the explicability of each topic, the number of topics in our case was set at 3.

5.1.2. Topic clusters

As depicted in Table 1, three topics have been identified. The proportions of Topics 1–3 in trajectories were 37.7, 35.3, and 27.0%, respectively. The top 12 keywords, most relevant to each topic, were selected and used to interpret the topic. Topic 1 is associated with recreational cycling and includes green urban areas, water courses, and port areas, reflecting diverse leisure activities. Topic 2 centers on the urban exploration undertaken by a city cyclist, with nearly all keywords intricately tied to urban amenities. Topic 3 focuses on green cycling for nature, as evidenced by the prevalence of keywords, such as forest, pasture, and arable land.

Considering trajectories and the corresponding assigned topics as input, we proceed to visually depict the feature of each topic through dimensionality reduction using PCA. As seen in Figure 10, the points appear roughly on top of each other, which is because WLDA allows a word to belong to several clusters with varying degrees rather than inducing only distinct clusters. Nevertheless, the three topics still exhibit a degree of clustering effects.



(a) Topics = 3. (b) Topics = 5.

Figure 9. PCoA results of the topic distribution. (a) Topics = 3. (b) Topics = 5.

Table 1. The latent topics of cycling trajectories.

Topics	Description	Keywords	Percentage
Topic 1	Leisure tour	Path, bench, sport and leisure facilities, green urban areas, watercourses, waste_basket, footway, pastures, discontinuous urban fabric, cycleway, service, port_areas	37.7%
Topic 2	City tour	Discontinuous urban fabric, continuous urban fabric, industrial or commercial units, secondary, residential, primary, tertiary, memorial, restaurant, road and rail networks and associated land, waste_basket, fast_food	35.3%
Topic 3	Green tour	Non-irrigated arable land, track, broad-leaved forest, discontinuous urban fabric, pastures, tourist_info, bench, unclassified, residential, tertiary, mixed forest, secondary	27.0%

5.2. WLDA-informed sampling and recommendation of trajectories

5.2.1. Spatial distribution of cycling topics

To further visually represent trajectories' spatial distribution and aggregation characteristics, we performed line density analysis on trajectories corresponding to each topic, as shown in Figure 11. Figure 11(a) shows the result of the overlay analysis for the topics. The spatial distribution of the hotspot areas for the three topics is different, and there is also variation in their respective coverage. As depicted in Figure 11(b), trajectories in this topic exhibit a limited spatial range, predominantly distributed along the Rhine River, and manifest a noticeable aggregation effect. We deduced that the cycling intention in this topic primarily entails short-distance leisure activities, which radiate to the surrounding green parks centered in the city. In contrast to the leisure tour, the aggregation of trajectories within the city tour is more distinctive but also has broader coverage, with the majority of trajectories located in the urban area of Bonn, as shown in Figure 11(c). The green tour boasts the most extensive coverage and manifests a distinct clustering effect within green spaces, as Figure 11(d) depicts. Diverging from the previous two topics, green cycling typically entails long-distance cycling, as evidenced by the fact that many of the trajectories also extend to forested areas beyond the city.

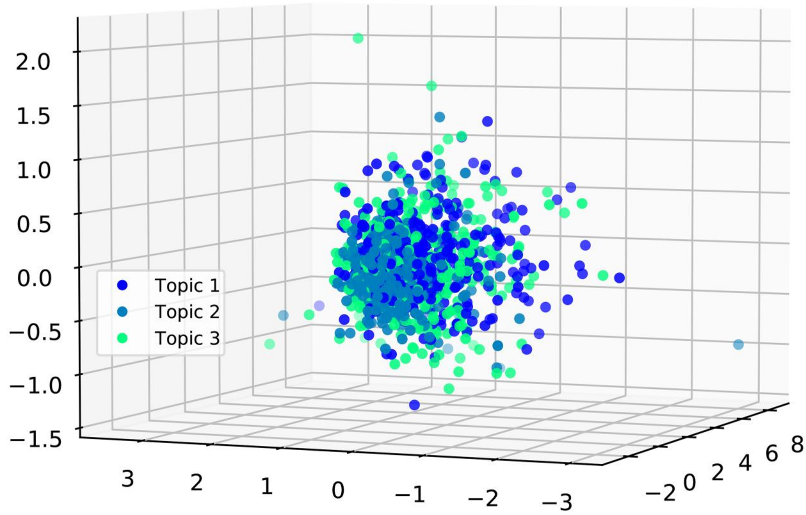


Figure 10. Feature representation of topics. Each point indicates the feature representation of an individual trajectory in three dimensions, comprising the most promising features determined by PCA ranking.

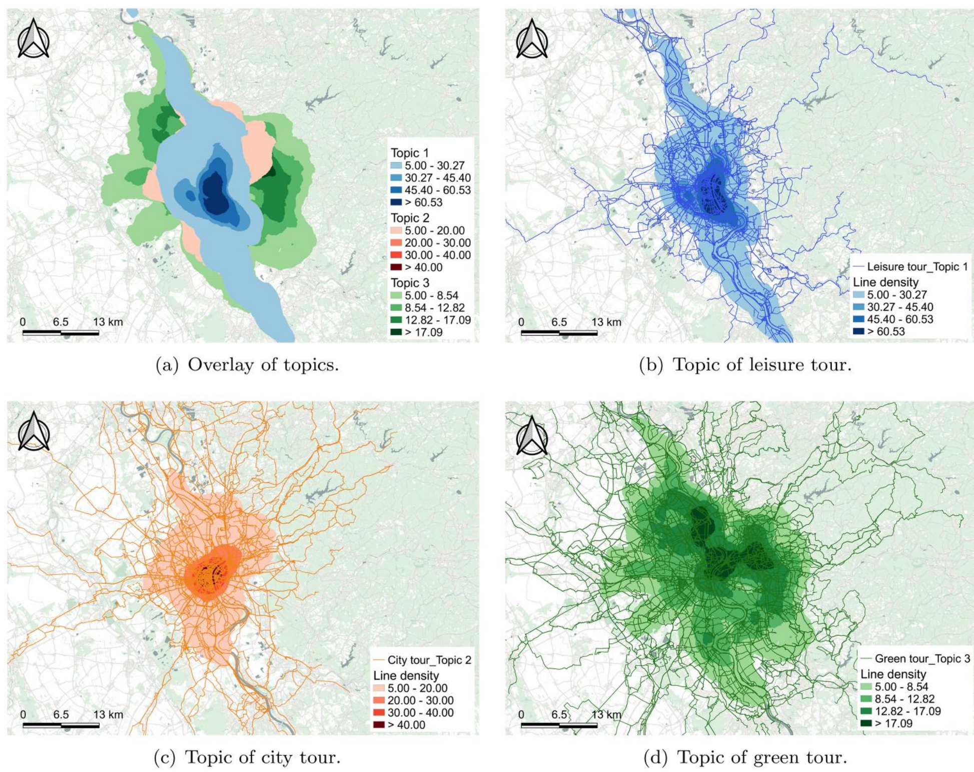


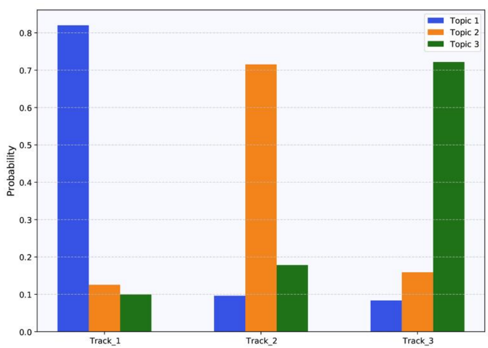
Figure 11. Line density analysis of trajectory topics and their spatial distribution. (a) Overlay of topics. (b) Topic of leisure tour. (c) Topic of city tour. (d) Topic of green tour.

5.2.2. Cycling routes recommendation based on topic probability

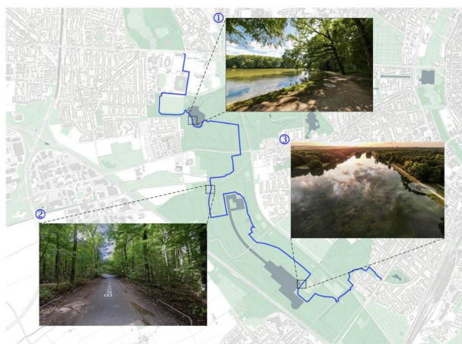
Building upon our WLDA model, we can further calculate the distance between the preference vector of cyclists and the track-topic probability distribution. By identifying the minimal distance, we can offer insightful suggestions to unfamiliar cyclists, as shown in Figure 12(a). Let us set up a scenario: a tourist pursuing an ideal leisurely experience in Bonn and seeking a ride with lush greenery near the city, the preference is roughly like leisure 82%, green 10%, and the city 8%. To meet these criteria, we could sample and recommend a trajectory based on the minimal ED, as depicted in Figure 12(b). This selected trajectory unfolds a picturesque route encompassing parks and tranquil lakes, perfectly tailored for a delightful short weekend leisure ride. In addition to the leisure routes, we have selected two specific routes from the other cycling topics for representation. Figure 12(c) shows an urban cycling route predominantly confined within city limits. Conversely, Figure 12(d) is a scenic long-distance cycling route situated away from urban areas, encompassing natural elements, such as forests, grasslands, and farmlands.

6. Discussion

In this article, we developed a WLDA model to recommend cycling routes from a semantic mining perspective, which offers novel research avenues for applying



(a) Topic probability of tracks.



(b) Track 1-leisure tour.

(c) Track 2-city tour.

(d) Track 3-green tour.

Figure 12. Topic probability-based trajectory recommendation and visualization. The above photo examples are from Google Street View. (a) Topic probability of tracks. (b) Track 1-leisure tour. (c) Track 2-city tour. (d) Track 3-green tour.

machine learning in route planning. However, there are still some points that need to be further discussed.

Regarding spatial context integration, following the statement of Meng and Zheng (2023), we only considered landuse data, POI information, and road type data due to the limited availability of data. However, we are trying to extract more information from the semantically rich and high-resolution CityGML to expand the corpus. In our case, the buffer radius used to integrate the spatial context was set at 50 meters. The aim is to intersect the potential landuses located behind buildings on the main street, but it can also capture features from adjacent streets. A more sensible solution would be to set the buffer radius contextually based on the class and width of the road and the type of amenities in a dynamic fashion.

In our case, the route recommendation is designed primarily for recreational cycling. We assume that cyclists care more about the spatial context along the route than the origin and destination. Therefore, our method can actively sample a cycling route that matches cyclists' preferences from a large dataset of trajectories. However, to encourage people to switch from cars to bicycles, the method should also be able to suggest appropriate routes for everyday journeys and consider users' pre-defined origins and destinations. To address this, our future work will deal with using Large Language Models (LLMs) (Zhu *et al*. 2024) to learn our topic-word distributions to sample new routes, thereby accommodating more flexible cycling preferences.

In terms of the evaluation of the results, we used the JSD value to evaluate the topic modeling results of cycling trajectories, but the recommended cycling routes need further quantitative validation, which is a priority for future work. One option is to recruit participants to cycle along our recommended routes and evaluate the recommendations based on their feedback, and this is undoubtedly time-consuming in practice. This option will be more feasible once our app for route recommendations for recreational cycling is released. Another option is to use Google Street View to evaluate the recommended route, but this approach makes it difficult to capture the real sense of what the cyclist experiences.

Currently, we focus more on the algorithm for retrieving and recommending cycling routes and even generating new routes by learning an underlying probability distribution. However, the output of this article may have broader applications. On the one hand, integrating spatial context with trajectories could enrich the data input for OSRM and CycleStreets, the popular bicycle routing and advocacy tools. On the other hand, combining the data-driven generation of weights files from this article with OSRM and CycleStreets profiles can customize individual cycling topics for cyclists, opening up more possibilities for OSRM and CycleStreets cycling recommendations.

7. Conclusion

Early studies on route recommendation focused on discovering the most efficient routes instead of the attractive ones. Subsequently, multiple criteria analysis and machine learning were used for route recommendation, and these approaches ignored the significance of spatial context, consequently leading to limited interpretability. From the authors' point of view, a comprehensive perspective on cycling route recommendation should not solely consider the trajectory itself but also incorporate the spatial context and underlying semantics of road segments. In this context, we proposed a novel method for recommending cycling routes leveraging WLDA with crowd-sourced trajectories. First, the spatial context was mapped into the trajectory. Subsequently, WLDA was employed to mine the latent topics in cycling trajectories. Finally, the insightful suggestions for unfamiliar cyclists were sampled, building upon the topic probability. We aim to extract hidden topics of cycling semantics to suggest cyclists choose the desirable cycling routes according to their preferences.

We conclude that it is possible to retrieve meaningful topics of cycling behavior from crowd-sourced trajectories. In particular, by integrating urban spatial context with trajectories to create a corpus for mining cycling route topics using NLP, we can sample and recommend routes based on topic probabilities that match users' preferences. In summary, this article incorporated a weighting mechanism into LDA, which allows for sampling and recommending cycling routes from a large dataset of crowd-sourced trajectories, successfully shifted the route recommendation paradigm from GIS analysis to a semantic mining perspective, yielding highly interpretable results and offering novel research avenues for applying machine learning in route planning. In the future, a topic worthy of investigation exploits the topic modeling results of trajectories. This valuable information could be coupled with LLMs or Mixed-Integer Linear Programming (MILP) to generate an entirely new trajectory for tourists in a new city. In addition, we will explore the possibility of integrating the weight files generated by our approach with OSRM and CycleStreets profiles to customize cycling topics for individual cyclists.

Notes

1. https://project-osrm.org/.
2. https://www.cyclestreets.net/.
3. https://dictionary.cambridge.org/dictionary/english/trajectory.
4. www.gpsies.com.
5. www.openstreetmap.org.
6. https://open.nrw/.

Acknowledgments

The authors would like to express their gratitude to the OpenStreetMap community and the Open North Rhine-Westphalia (NRW) for their invaluable contribution to open geospatial data, as well as to the editor and anonymous reviewers for their careful reading and constructive suggestions.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This paper was supported by the National Natural Science Foundation of China (Grant No. 42201446), the Chinese Postdoctoral Science Foundation (Grant No. 2024T170742), Postdoctoral Fellowship Program of CPSF (Grant No. GZC20232185), and the Fundamental Research Funds for the Central Universities (Grant No. 2682024CX095).

Notes on contributors

Weilian Li is a postdoctoral research associate in the Computational Methods Lab at the HafenCity University Hamburg. He received his Ph.D. in Geomatics Science and Technology at Southwest Jiaotong University in December 2020. Then, he finished his postdoc research stay in the Geoinformation Group at the University of Bonn in October 2023. He has a strong interest in topics related to smart cities, disaster management, and digital twins. He proposed the idea and contributed to this article's methodology, implementation, experiment, and writing.

Jan-Henrik Haunert holds a diploma and doctoral degree in geodesy and geoinformatics from the University of Hannover. He has been a postdoctoral researcher at the Institute of Computer Science at the University of Wurzburg and a professor of geoinformatics at the University of Osnabruck. In 2016, he took up a full professorship for geoinformation at the University of Bonn. His research is concerned with the development of efficient algorithms for geovisualization and spatial analysis. He contributed to the article's methodology, review, and proofreading.

Axel Forsch finished his Master of Science in the Institute of Geodesy and Geoinformation at the University of Bonn in 2019. Currently, he continues his studies at the University of Bonn within the frame of a PhD program regarding Volunteered Geographic Information. His research focuses on the algorithmic analysis and visualization of movement patterns. He proposed the idea and contributed to the experimental data processing.

Jun Zhu is a professor and the deputy dean of the Faculty of Geosciences and Engineering at Southwest Jiaotong University. He received his Ph.D. in Cartography and GIS from the Chinese Academy of Sciences. His research interests include computer vision, 3D GIS, and virtual geographic environments. He contributed to the literature review and proofreading.

Qing Zhu is a professor and the director of the Research Committee in the Faculty of Geosciences and Engineering at Southwest Jiaotong University. He received his Ph.D. in railway engineering from North Jiaotong University. His research interests include photogrammetry, geographic information systems, and virtual geographic environments. He played a crucial role in review and proofreading.

Youness Dehbi is a full professor and the Chair of the Computational Methods Lab at the HafenCity University Hamburg. He received his Ph.D. degree in the Institute of Geodesy and Geoinformation at the University of Bonn in 2016. His research focuses on urban reasoning, analytics, pattern recognition, machine learning and statistical relational learning for geoinformation. He proposed the idea and contributed to this article's methodology, writing, and review.

Data and codes availability statement

The data and codes that support the findings of this study can be found online at https://code.computationalmethods.hcu-hamburg.de/papers/cycling_trajectories_wlda.

References

Abdullah, M., *et al.*, 2021. Measuring changes in travel behavior pattern due to COVID-19 in a developing country: A case study of Pakistan. *Transport Policy*, 108, 21–33.

AlAgha, I., 2021. Topic modeling and sentiment analysis of Twitter discussions on COVID-19 from spatial and temporal perspectives. *Journal of Information Science Theory & Practice*, 9, 35–53.

Behr, T., *et al.*, 2021. Map matching for semi-restricted trajectories. In: K. Janowicz and J.A. Verstegen, eds. *11th International Conference on Geographic Information Science (GIScience 2021) – Part II, Schloss Dagstuhl – Leibniz-Zentrum für Informatik*, Dagstuhl, Germany, 12:1–12:

16. Available from: https://drops-dev.dagstuhl.de/entities/document/10.4230/LIPIcs.GIScience.2021.II.12.

Biassoni, F., *et al.*, 2023. Choosing the bicycle as a mode of transportation, the influence of infrastructure perception, travel satisfaction and pro-environmental attitude, the case of Milan. *Sustainability*, 15 (16), 12117.

Blei, D.M., Ng, A.Y., and Jordan, M.I., 2003. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.

Brauer, A., Mäkinen, V., and Oksanen, J., 2021. Characterizing cycling traffic fluency using big mobile activity tracking data. *Computers, Environment and Urban Systems*, 85, 101553.

Buehler, R. and Pucher, J., 2021. COVID-19 impacts on cycling, 2019–2020. *Transport Reviews*, 41 (4), 393–400.

Chen, C.F. and Chen, P.C., 2013. Estimating recreational cyclists' preferences for bicycle routes–evidence from Taiwan. *Transport Policy*, 26, 23–30.

Chen, W., *et al.*, 2018. Social functional mapping of urban green space using remote sensing and social sensing data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 146, 436–452.

Chu, D., *et al.*, 2014. Visualizing hidden themes of taxi movement with semantic transformation. 2014 IEEE Pacific Visualization Symposium. IEEE, 137–144.

Cui, G., Luo, J., and Wang, X., 2018. Personalized travel route recommendation using collaborative filtering based on GPS trajectories. *International Journal of Digital Earth*, 11 (3), 284–307.

de Oliveira e Silva, R.A., *et al.*, 2022. Personalized route recommendation through historical travel behavior analysis. *GeoInformatica*, 26 (3), 505–540.

Đerek, J. and Sikora, M., 2019. Bicycle route planning using multiple criteria GIS analysis. 2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM). IEEE, 1–5.

Desjardins, E., *et al.*, 2022. Correlates of bicycling trip flows in Hamilton, Ontario: fastest, quietest, or balanced routes? *Transportation*, 49 (3), 867–895.

Ding, X., Fan, H., and Gong, J., 2021. Towards generating network of bikeways from Mapillary data. *Computers, Environment and Urban Systems*, 88, 101632.

Dingil, A.E. and Esztergár-Kiss, D., 2021. The influence of the COVID-19 pandemic on mobility patterns: the first wave's results. *Transportation Letters*, 13 (5–6), 434–446.

Eren, E. and Uz, V.E., 2020. A review on bike-sharing: the factors affecting bike-sharing demand. *Sustainable Cities and Society*, 54, 101882.

Etminani-Ghasrodashti, R., Paydar, M., and Ardeshiri, A., 2018. Recreational cycling in a coastal city: investigating lifestyle, attitudes and built environment in cycling behavior. *Sustainable Cities and Society*, 39, 241–251.

Fang, F., *et al.*, 2022. Spatial context-aware method for urban land use classification using street view images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 192, 1–12.

Feng, Y., Brenner, C., and Sester, M., 2020. Flood severity mapping from volunteered geographic information by interpreting water level from images containing people: a case study of hurricane Harvey. *ISPRS Journal of Photogrammetry and Remote Sensing*, 169, 301–319.

Forsch, A., *et al.*, 2023. Inferring routing preferences from user-generated trajectories using a compression criterion. *Journal of Spatial Information Science*, 26, 99–124.

Fraser, S.D. and Lock, K., 2011. Cycling for transport and public health: a systematic review of the effect of the environment on cycling. *European Journal of Public Health*, 21 (6), 738–743.

Gao, X., *et al.*, 2023. Visual exploration of cycling semantics with GPS trajectory data. *Applied Sciences*, 13 (4), 2748.

Gelfand, A.E., 2000. Gibbs sampling. *Journal of the American Statistical Association*, 95 (452), 1300–1304.

Goodchild, M.F., 2007. Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69 (4), 211–221.

Guo, Y., *et al.*, 2021. Dockless bike-sharing as a feeder mode of metro commute? The role of the feeder-related built environment: analytical framework and empirical evidence. *Sustainable Cities and Society*, 65, 102594.

Han, X., *et al.*, 2020. Using social media to mine and analyze public opinion related to COVID-19 in China. *International Journal of Environmental Research and Public Health*, 17 (8), 2788.

Harms, L., Bertolini, L., and Te Brömmelstroet, M., 2014. Spatial and social variations in cycling patterns in a mature cycling country exploring differences and trends. *Journal of Transport & Health*, 1 (4), 232–242.

Hu, S., *et al.*, 2021. Examining spatiotemporal changing patterns of bike-sharing usage during COVID-19 pandemic. *Journal of Transport Geography*, 91, 102997.

Huang, W., *et al.*, 2015. Predicting human mobility with activity changes. *International Journal of Geographical Information Science*, 29 (9), 1569–1587.

Imran, M., *et al.*, 2015. Processing social media messages in mass emergency: a survey. *ACM Computing Surveys*, 47 (4), 1–38.

Li, W., *et al.*, 2021. An augmented representation method of debris flow scenes to improve public perception. *International Journal of Geographical Information Science*, 35 (8), 1521–1544.

Li, W., *et al.*, 2023. Social media insights on public perception and sentiment during and after disasters: the European floods in 2021 as a case study. *Transactions in GIS*, 27 (6), 1766–1793.

Li, W., *et al.*, 2024a. A 3D virtual geographic environment for flood representation towards risk communication. *International Journal of Applied Earth Observation and Geoinformation*, 128, 103757.

Li, W., *et al.*, 2024b. Visual attention-guided augmented representation of geographic scenes: a case of bridge stress visualization. *International Journal of Geographical Information Science*, 38 (3), 527–549.

Liu, H., *et al.*, 2019. Visual analytics of taxi trajectory data via topical sub-trajectories. *Visual Informatics*, 3 (3), 140–149.

Lovelace, R., Félix, R., and Carlino, D., 2022. Exploring jittering and routing options for converting origin-destination data into route networks: towards accurate estimates of movement at the street level. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-4/W1-2022, 279–286.

Mahfouz, H., Lovelace, R., and Arcaute, E., 2023. A road segment prioritization approach for cycling infrastructure. *Journal of Transport Geography*, 113, 103715.

Martín, B. and Páez, A., 2019. Individual and geographic variations in the propensity to travel by active modes in Vitoria-Gasteiz, Spain. *Journal of Transport Geography*, 76, 103–113.

Meng, S. and Zheng, H., 2023. A personalized bikeability-based cycling route recommendation method with machine learning. *International Journal of Applied Earth Observation and Geoinformation*, 121, 103373.

Mou, N., *et al.*, 2022. Personalized tourist route recommendation model with a trajectory understanding via neural networks. *International Journal of Digital Earth*, 15 (1), 1738–1759.

Nadi, S. and Delavar, M.R., 2011. Multi-criteria, personalized route planning using quantifier-guided ordered weighted averaging operators. *International Journal of Applied Earth Observation and Geoinformation*, 13 (3), 322–335.

Nasution, S.M., *et al.*, 2022. Personalized route recommendation using F-AHP-express. *Sustainability*, 14 (17), 10831.

Neishabouri, A. and Desmarais, M.C., 2020. Reliability of perplexity to find number of latent topics. The Thirty-Third International Flairs Conference.

Nguyen, M.H. and Pojani, D., 2022. The emergence of recreational cycling in Hanoi during the COVID-19 pandemic. *Journal of Transport & Health*, 24, 101332.

Nugroho, R., *et al.*, 2020. A survey of recent methods on deriving topics from Twitter: algorithm to evaluation. *Knowledge and Information Systems*, 62 (7), 2485–2519.

Oehrlein, J., *et al.*, 2018. Inferring routing preferences of bicyclists from sparse sets of trajectories. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-4/W7, 107–114.

Ogilvie, D., *et al.*, 2011. An applied ecological framework for evaluating infrastructure to promote walking and cycling: the iconnect study. *American Journal of Public Health*, 101 (3), 473–481.

Pucher, J. and Buehler, R., 2017. Cycling towards a more sustainable transport future. *Transport Reviews*, 37 (6), 689–694.

Siriaraya, P., *et al.*, 2020. Beyond the shortest route: a survey on quality-aware route navigation for pedestrians. *IEEE Access*, 8, 135569–135590.

Sultan, J., *et al.*, 2017. Extracting spatial patterns in bicycle routes from crowdsourced data. *Transactions in GIS*, 21 (6), 1321–1340.

Tang, Y., *et al.*, 2018. Visual analysis of traffic data based on topic modeling (chinaVis 2017). *Journal of Visualization*, 21 (4), 661–680.

Teixeira, J.F., Silva, C., and Moura e Sá, F., 2021. Empirical evidence on the impacts of bikesharing: a literature review. *Transport Reviews*, 41 (3), 329–351.

Wang, Z. and Ye, X., 2018. Social media analytics for natural disaster management. *International Journal of Geographical Information Science*, 32 (1), 49–72.

Yan, Y., *et al.*, 2020. Volunteered geographic information research in the first decade: a narrative review of selected journal articles in GIScience. *International Journal of Geographical Information Science*, 34 (9), 1765–1791.

Yang, L., *et al.*, 2020. Accessibility to transit, by transit, and property prices: spatially varying relationships. *Transportation Research Part D: Transport and Environment*, 85, 102387.

Yang, L., *et al.*, 2021. To walk or not to walk? examining non-linear effects of streetscape greenery on walking propensity of older adults. *Journal of Transport Geography*, 94, 103099.

Zhang, T., *et al.*, 2021. A topic model based framework for identifying the distribution of demand for relief supplies using social media data. *International Journal of Geographical Information Science*, 35 (11), 2216–2237.

Zhu, S., 2022. Multi-objective route planning problem for cycle-tourists. *Transportation Letters*, 14 (3), 298–306.

Zhu, J., *et al.*, 2024. A flood knowledge-constrained large language model interactable with GIS: enhancing public risk perception of floods. *International Journal of Geographical Information Science*, 38 (4), 603–625.