Spatial keyword search: a survey

Lisi Chen[1] · Shuo Shang[1] · Chengcheng Yang[2] · Jing Li[2]

Abstract

Spatial keyword search has been playing an indispensable role in personalized route recommendation and geo-textual information retrieval. In this light, we conduct a survey on existing studies of spatial keyword search. We categorize existing works of spatial keyword search based on the types of their input data, output results, and methodologies. For each category, we summarize their common features in terms of input data, output result, indexing scheme, and search algorithms. In addition, we provide detailed description regarding each study of spatial keyword search. This survey summarizes the findings of existing spatial keyword search studies, thus uncovering new insights that may guide software engineers as well as further research.

1 Introduction

Location-based multimodal data, that most ubiquitous of location-based services and geo-tagged web contents, has been keeping increasing at a very high speed over the last decade. It allows both users and service providers to generate and disseminate geo-located textual content through a variety of social media and map services (e.g., micro-blogging platforms, photo sharing platforms, peer-to-peer ridesharing apps, foot delivery apps, and news websites). For example, Twitter, a popular micro-blogging service allowing mobile users to

✉ Shuo Shang
jedi.shang@gmail.com

Lisi Chen
lchen012@e.ntu.edu.sg

Chengcheng Yang
chengcheng.yang@inceptioniai.org

Jing Li
jing.li@inceptioniai.org

[1] UESTC, Chengdu, China

[2] Inception Institute of Artificial Intelligence, Abu Dhabi, United Arab Emirates

compose geo-tagged tweets with short text and locations (coordinates or semantic locations), has gained increased attention among researchers in different disciplines. Other examples include gao-tagged photos from social photo sharing websites (e.g., Flickr and Instagram), check-ins in location based social networks (e.g., Foursquare), reviews on local business websites (e.g., Yelp and TripAdvisor) that contain both text content and locations of points of interest (POIs), online local or regional news with long documents and location tags, and trajectories from peer-to-peer ridesharing apps (e.g., Uber, Grab Taxi) and food delivery apps (e.g., Uber Eats, Delivery.com, Grubhub).

According the input data type, location-based multimodal data can be classified into two categories: standalone location-based multimodal data (geo-textual object) and connected location-based multimodal data (e.g., keyword-based route, trajectory with category information). One representative example of standalone location-based multimodal data is POIs generated from Foursquare, Yelp, and Google Maps. Many POIs are being associated with category information (e.g., Restaurant, Hotel, Bar, etc.) or users' descriptions and reviews. As for connected location-based textual data, popular examples include routes associated with keywords and trajectories associated with category and time information. According to arrival pattern, we can classify location-based multimodal data into static location-based multimodal data and streaming location-based multimodal data. Static location-based multimodal data can be regarded as a collection of objects. We basically model it as a data collection with very infrequent update. Whilst streaming location-based multimodal data can be modeled as a stream of continuously arriving objects, which is infeasible for us to index all of them. The most common example of streaming location-based textual data is geo-tagged tweets.

Location-based multimodal data from various sources has been characterized by big volume and multi-dimensions. According to existing studies [1, 2], the volume of location-based multimodal data is keeping growing rapidly over the last couple of decades. Besides the textual dimension and geo-spatial dimension, location-based multimodal data can be featured with other dimensions, including time dimension (i.e., the creation time of an item), user dimension (i.e., user who posted the item), dimension of category (i.e., category information of an item).

In this survey, we investigate existing studies on querying for location-based multimodal data on road networks. In particular, we classify existing literature based on the types of their input data, output results, and methodologies. For each category, we summarize their common features in terms of input data, output result, indexing scheme, and search algorithms. Description regarding each study of location-based multimodal data management is presented.

2 Related work

This section introduces existing survey and evaluation studies on the topic of location-based data management.

2.1 Survey of location-based data management

The earliest survey of spatial keyword search is conducted by Cao et al. [3]. It reviews some basic spatial keyword queries (standard spatial keyword queries, collective spatial keyword queries, and moving spatial keyword queries) that were proposed before 2012. However, this survey does not include a host of recent studies related to spatial keyword search.

Bao et al. [4] present a survey of recommendation studies on location-based social netowrks. Specifically, it analyzes the data formulation, the methodology of recommendation, and the objectives of various recommendation tasks. This survey mainly focuses on recommendation problem instead of querying and data management. Kong et al. [5] and Feng et al. [6] introduce the trajectory data management and mining from the perspective of applications and services. Mahmood [7] present an overview and a broad classification of existing spatio-temporal access techniques published between 2010 and 2017. It covers a wide range of geo-textual indexing structures. A book chapter is published on scalable processing of spatial keyword queries [8]. It summarizes the various types of queries that execute over geo-textual data. However, it does not cover the queries executing over trajectory data. Cong et al. [9] present a tutorial that provides an overview of the problems addressed in the area of spatial keyword querying and suggests some important open problems and new research direction. A high-level review of studies on searching and mining geo-textual data for exploration before 2016 is presented [10].

2.2 Experiment and evaluation of location-based data management

Chen et al. [11] presents a survey of 12 basic geo-textual indexing structures proposed before 2013. They develop a benchmark that enables the comparison of the spatial keyword query performance on different indexing structures. Note that they only consider three types of basic spatial keyword query: Boolean-Boolean Query, Boolean-Ranking Query, and Full-Ranking Query (cf. Section 4.1). Liu et al. [12] provide an evaluation of 12 state-of-the-art POI recommendation models, which uncover a number of significant results regarding the utilization of POI recommendation models in a variety of scenarios.

3 Data formulation

This section investigates the classification of locaion-based multimodal data defined by existing studies. Most of the related work in this area study the problem of processing spatial keyword queries over geo-textual data. We proceed to introduce the data space (within which the data are positioned), the location-based multimodal data, and the query formulation, respectively.

3.1 Location-based multimodal data

Location-based multimodal data is the items that are being searched or queried. Each item is assumed to contain a location and a set of attributes of different types. We proceed to present a brief introduction of each common attribute.

Location attribute: The location attribute can be presented in the following forms:

(1) *Physical point in a 2D space*: The point specifying a location is represented by a tuple with latitude and longitude (e.g., 23°25′43″ N, 54°21′78″ W);

(2) *Physical region in a 2D space*: GPS location may be inaccurate due to various reasons (e.g., incorrect settings in end device, weak signal). As a result, we may also regard the location as a region (most often a circular region) on a 2D space.

(3) *Semantic location*: Basically, semantic locations are organized by a GIS dictionary (i.e., geo-spatial definition glossary), which is a hierarchical structure that indexes geographical terms (e.g., Ann Arbor, MI, U.S.A.).

(4) *Location-based terms*: Some documents (e.g., news, reports) do not have a specific
field to specify their locations. However, we may extract location-based terms (e.g.,
football club in north London) to figure out their most-likely locations.

Text attribute: Most of the existing work assumes that the text attribute of a geo-textual
object is a simple, unstructured text string, which is modeled as a term vector or just a
set of keywords.

Time attribute: Location-based multimodal items may contain temporal information. It
may be a timestamp (e.g., 11:59 P.M., Jun 15, 2019), capturing the time when the item
was created (e.g., timestamp of a tweet or review), a time range (e.g., Aug 3, 2019 – Aug
7, 2019), representing the date of a particular event, or multiple time ranges (e.g., 9:30
A.M. to 5:00 P.M. from Monday to Friday), representing the opening hours of a POI.

Category attribute: We may assign one or more categories to a location-based multi-
modal items. Category attributes may denote the classification of an item (e.g., Cafe,
restaurant, hotel, cinema), review summarization (e.g., cosy, sea view, long waiting time),
etc.

3.2 Distance metrics

We introduce distance metrics defined by existing studies for measuring the spatial distance
between two items. The underlying space used by existing studies on location-based multi-
modal querying can be classified into two categories: Euclidean space and spatial-network
space. We present how to compute the spatial distance between two items in Euclidean
space and spatial-network space, respectively.

Euclidean space: Given two items i_1 and i_2, the spatial distance between i_1 and i_2 on
Euclidean space can be regarded as the great-circle distance between i_1 and i_2 on the
earth.

Spatial-network space: As for spatial-network space, items are located at a spatial net-
work, which is a directed or undirected graph where each vertex has a coordinate (i.e.,
latitude and longitude) and each edge has a weight that captures the great-circle distance
between the two end-points on the edge. Note that in the context of spatial networks,
travel time may also be considered as the relevant distance notion. In that case, travel
times, sometimes time varying, are associated with edges.

4 Querying for individual location-based multimodal data

This section introduces existing studies of query processing over a collection of individual
location-based multimodal data or a stream of individual location-based multimodal data.

4.1 Basic spatial keyword queries

Given a collection of location-based multimodal items, queries that returns a subset of indi-
vidual items are named as *item-wise spatial keyword queries*. The main characteristics of an
item-wise spatial keyword query q are summarized as follows:

(1) q contains both spatial and textual arguments;
(2) The result items returned by q are independent of each other.

We proceed to present the spatial argument and textual argument in detail.

Spatial argument: A spatial argument is basically a spatial point that denotes the location of the query issuance. However, the spatial argument may also be a region or multiple regions. The region(s) serve as a filter that filters out items whose locations do not fall in (one of) the region(s).

Textual argument: The textual argument (or query keywords) may be of the same form as the textual attributes of items. In general, the textual attribute is a set of keywords provided by a user. The keywords may either serve as a filter that filters out items whose textual attributes do not contain the query keywords or serve as arguments of a particular textual similarity function.

Item-wise spatial keyword queries are quite simple. However, these queries lay a solid foundation for various complex spatial keyword queries of different objectives. The definitions and processing of item-wise spatial keyword queries are extensively investigated by Chen et al. [11]. Here we summarize the survey result of basic item-wise spatial keyword queries presented by Chen et al. [11].

Let D be a set of location-based multimodal items. Each item $i \in D$ is defined as a pair $(i.loc, i.doc)$, where $i.loc$ is a spatial point and $l.doc$ is a text document. We proceed to present three types of basic item-wise spatial keyword queries.

Boolean-Boolean (BB) Query [13–19] Given a BB query $q = \langle r, w \rangle$, where r denotes the query region (i.e., most often a circular or rectangular spatial region) and w is a set of keywords. Query q returns a subset of D, denoted by D_q, which contains items s.t. $\forall i \in D_q (dist(i.loc \in q.r \wedge q.w \subseteq i.doc)$. Specifically, D_q denotes items in D whose documents contain all the query keywords ($q.w$) and locations fall in $q.r$.

Boolean-Ranking (BR) Query [19–24] A BR query $q = \langle p, w, k \rangle$ has three elements, where p is a spatial point (query location), w is a set of query keywords, and k represents the number of items to returned by q. The result of q, D_q, is a k-subset of D s.t. the document of each item i in D_q contains all the keywords in $q.w$. and i is a k-NN (i.e., k nearest neighbors) of q.

Full-Ranking (FR) Query [19, 24–27] An FR query $q = \langle p, w, k, \alpha \rangle$ has four elements, where p is a spatial point (query location), w is a set of query keywords, k is the number of items to returned by q, and $q.\alpha$ is a preference parameter that balances the weight between textual relevancy and spatial proximity. An FR query q returns k items ranked according to a relevance score that takes both spatial proximity and text relevancy into account. In particular, the relevance between an item i and an FR query q is defined by Eq. 1

$$Rel(i, q) = F(SRel(i, q), TRel(i, q)), \tag{1}$$

where $SRel(i, q)$ denotes the *spatial relevance* between i and q, which is calculated as follows:

$$SRel(i, q) = G(dist(i.loc, q.p)),$$

where $G(x)$ is a monotone decreasing function ranging between 0 and 1 and $dist(i.loc, q.p)$ is the Euclidean distance between i and q. Notation $TRel(i, q)$ denotes the textual relevancy between $i.doc$ and $q.w$. Function $F(x, y)$, ranging between 0 and 1, is a monotone increasing function in both x and y.

We may compute $TRel(i, q)$ by using an information retrieval model, such as the language model (e.g., [25]), cosine similarity (e.g., [27]), or BM25 (e.g., [18]) normalized to a scale similar to the spatial relevance part.

4.2 Advanced spatial keyword queries

We proceed to introduce advance spatial keyword queries, which incorporate additional elements to basic spatial keyword queries.

Reverse spatial keyword search Existing studies on reverse spatial keyword query can be classified into two categories: Reverse FR query and Reverse BR query. Both reverse FR query and reverse BR query regard "item" as "query". In particular, a reverse BR query q returns all items whose BR query results include q. Likewise, a reverse FR query q returns items whose FR query results include q. In Euclidean space, reverse BR search is investigated by Fang et al. [28] and the reverse FR search is studied by Lu et al. [29–31]. In spatial network, Gao et al. [32] proposed a mechanism to process reverse BR search and Luo et al. [33] studies the problem of processing reverse FR query. Additionally, Shang et al. [34] investigate Reverse Path Nearest Neighbor (R-PNN) search that finds the most accessible locations in road networks. Given a route dataset and a query location set, it finds locations with the highest influence factor. Specifically, if a location o is the Path Nearest Neighbor (PNN) of k routes, the influence factor of o is defined as k and the R-PNN query retrieves the location that has the highest influence factor. The R-PNN query is an extension of the conventional reverse nearest neighbor search.

Collective spatial keyword search Collective spatial keyword query [35–39] basically consists of a location and a set of keywords. The query can be defined in various format. However, the basic aim of collective spatial keyword query is to retrieve a set of items such that: (1) the keywords of result items cover the query keywords; (2) the items are expected to be close to the query location; (3) The result item set has the smallest inter-item distances. Specifically, Requirement (2) is optional (the m-Closest Keywords (mCK) query [39, 40] does not have Requirement (2)). Requirement (3) can be defined by various objective functions (i.e., Sum function [35, 36], Max-Max function [35–37], and Min-Max function [35]). The processing of collective spatial keyword query is NP-hard. Existing studies mainly focus on developing efficient exact search algorithm, greedy-based search algorithm, and approximate search algorithm with provable approximation bounds, to answer the queries.

Diversity-aware spatial keyword search Diversity-aware spatial keyword search not only considers the relevancy between query and items, but also takes the dissimilarity between items into account. Mehta et al. [41] defines spatial-temporal-keyword query that combines keyword search and the problem of maximizing the spatio-temporal coverage and diversity of the top-k items returned by the query.

The basic query result diversification techniques lay foundation for the study of diversity-aware spatial keyword search. In particular, Drosou et al. [42] study the problem of maintaining the k-set with the highest diversity score over a sliding window on the basis of the Max-min diversification metric. A cover-tree index is maintained for selecting the most diverse k-set on the fly over a sliding window. Minack et al. [43] develop a streaming-based query result diversification mechanism to find a diverse set based on both Max-sum and Max-min diversification metrics. Liang et al. [44] propose a streaming diversification algorithm that combines information acquired by dynamic Dirichlet multinomial mixture topic model and a Proportionality based diversification scheme. Cheng et al. [45] focus on the problem of selecting the minimum number pf representative subset of items from social media for a small group of queries issued by the same user, which can be easily extended to support location-based multimodal data (e.g., geo-tagged tweets). Chen et al. [46] develop

a publish/subscribe system to support diversity-aware publish/subscribe. Zhang et al. [47] study spatial keyword search diversification on road networks.

Context-based spatial keyword search The problem of context-based spatial keyword search is to find a set of spatio-textual relevant items by additionally considering (1) semantic meaning or (2) textual information of collection of items when executing the spatial keyword search.

Derczynski et al. [48] conduct a survey regarding context-based querying, which includes longitudinal analyses on social media archives, local intent search, and spatio-temporal intent search. Qu et al. [49] define locality-based spatial keyword query that returns top-k sets of items based on a ranking function combining spatial proximity and textual similarity. The problem of processing locality-based spatial keyword query is NP-hard, and they develop exact search algorithm and greedy-based approximate search algorithm to process the query. Sun et al. [50] focus on interactive spatial keyword querying with semantics, which improve the basic spatial keyword queries by (1) making sense of the query keywords and (2) refining the understanding of query semantics through user interactions. They develop an interactive strategy to infer the latent query semantics of a given probabilistic topic model by taking user feedbacks as input. In each user-feedback iteration, the returned items are selected to ensure further precise inference of query semantics. Qian et al. [51] define a semantic-based spatial keyword query that returns the k items that are most similar to the query by considering spatial proximity and text semantic relevance. The text semantic relevance is measured by the coherence of semantic meanings between query and items. They develop a hybrid indexing structure that integrates spatial, textual, and semantic information in a hierarchical manner. Qu et al. [52] define a spatial keyword query that considers the spatial proximity and properties of the items. The query takes a result cardinality, a spatial range, and property-related preferences as parameters, and it finds a set of items with the given cardinality and in the given range that satisfies the preferences.

Besides, Han et al. [53–55] study the problem of mining neighborhood patterns from large labeled graph. These studies are applicable to context-based spatial keyword search. Yang et al. [56, 57] aim to solve the problem of spatial keyword search with fuzzy token matching, which considers approximate keyword matching rather than exact keyword matching.

POI recommendations The problem of Point-of-interests (POI) recommendations has been extensively studied in recent years. Specifically, the input of POI recommendation problem is as follows:

(1) A set of POIs P;
(2) A set of users U;
(3) A POI set associated with each user $u \in U$;
(4) Other sources regarding the factors related to the recommendation results (e.g., user locations, temporal information, social networks).

The output of the POI recommendation problem is recommending for each user a set of new POIs that are likely to be visited by the user. Existing studies on POI recommendation problem have various problem settings and different recommendation models are applied for processing evaluation data. Liu et al. [12] conduct an extensive evaluation that evaluates the performance of existing state-of-the-art POI recommendation models. Zhao et al. [58] conduct a survey of POI recommendation problem on location-based social networks.

Wang et al. [59] is one of the pioneering works on POI recommendation in location-based social networks (LBSNs). While it is evidenced that both friendships and geographical distances affect the users' behaviors, they proposed a random walk approach to produce improved performance. Instead of further refining recommendation models, Lu et al. [60] instead proposed a learning framework to integrate multiple models so that any individual user's personal preferences can be captured and tracked over time.

According to the classification result [12], existing models for POI recommendation can be categorized by Collaborative Filtering Model, Matrix Factorization Model that decomposes check-in matrix into user matrix and POI matrix, Poisson Factor Model that factorizes the user-POI check-in matrix by Poisson distribution, Link-based Model that uses a graph to model the relationship between users, and Hybrid Model that is regarded as a combination of two or more basic models.

Ye et al. [61] and Zhang et al. [62, 63] propose a collaborative-filtering-based POI recommendation model that takes user preference, location, and/or social network as the input. Cheng et al. [64] and Liu et al. [65] develop a POI recommendation model based on Poisson factor. Both of them take user preferences and location as input. Liu et al. [66], Lian et al. [67], and Li et al. [68] use matrix factorization model to solve the POI recommendation problem. They also take user preferences and location as input. Li et al. [69] additionally consider user social network to be the input. Gao et al. [70] propose an LRT model, which is based on matrix factorization model as well. Different from the models proposed by Liu et al. [66], Lian et al. [67], and Li et al. [68], the LRT model only takes user preferences and temporal information as input. Ma et al. [71] propose an auto-encoder model to learn the complex user-POI relations, which include a self-attentive encoder and a neighbor-aware decoder.

Spatial keyword query authentications Authentications of spatial keyword query is extensively studied in recent years. It aims at providing users with correct and accurate results for each spatial keyword query. Su et al. [72] is the first to study the problem of authenticating top-k spatial keyword queries in outsourced databases. Wu et al. [73] focus on a different type of query - moving top-k spatial keyword queries. Recently, Obiri et al. [74] study the problem of authenticating multiple spatial keyword queries simultaneously. Yue et al. [75] develop a revocable group signature scheme to provide general privacy-preserving authentications. Its technique is applicable for handling spatial keyword queries.

Privacy-preserving spatial keyword query The problem of privacy-preserving spatial keyword query is to process spatial keyword query without disclosing the information from the users who issue the query. Su et al. [76, 77] focus on privacy-preserving top-k spatial keyword queries over outsourced databases and untrusted cloud environments. Cui et al. [78] aim at processing Boolean-Boolean spatial keyword queries while preserving the confidential information from users.

Continuous spatial keyword query processing Most existing studies on processing continuous spatial keyword queries are based on the publish/subscribe framework, which is named as *location-based publish/subscribe (LPS)*. In particular, the publish/subscribe framework consists of two inputs: (1) publish items; (2) subscriptions. Here, in the context of location-based publish/subscribe, the publish items can be modeled as a stream of location-based multimodal items and subscriptions can be viewed as a set of continuous spatial keyword queries. A user may register a personalized subscription, which continuously receive publish items that satisfy the subscription requirements. Here, the subscription

requirements contain spatial requirement (e.g., an item should be located within a subscription region), textual requirement (e.g., an item should contains subscription keywords), and/or temporal requirement (e.g., an item should be published within a time frame).

Existing studies can be classified into three categories: (1) Item-based LPS [79–92]; (2) Frequency-based LPS [93–98]; (3) Cluster-based LPS [99–101]. Specifically, the idea of item-based LPS is to deliver relevant *items* to each subscription. Differently, the idea of frequency-based LPS is to deliver frequent *elements* to each subscription. Here, element generally denotes term (keyword) in an item. Cluster-based LPS delivers summarized information to each subscription (e.g., events, topics, and general clusters).

5 Search over connected location-based multimodal data

This section presents existing studies on querying over connected location-based multimodal data.

5.1 General trajectory search

5.1.1 Trajectory data

The continued proliferation of GPS-equipped mobile devices (e.g., vehicle navigation systems and smart phones) and the proliferation of online map-based services (e.g., Bing Maps,[1] Google Maps,[2] and MapQuest[3]s) enable the collection and sharing of trajectories. For example, the sites Bikely,[4] GPS-way-points,[5] Share-my-routes,[6] and Microsoft Geolife[7] enable such sharing, and more and more social network sites, including Twitter,[8] Facebook,[9] and Foursquare,[10] are starting to support trajectory sharing and search. This development motivates new studies of the management and analysis of massive trajectory data [102]. Spatial-keyword trajectory search is useful in many popular applications including route planning and recommendation, ridesharing, friend recommendation in social networks, and location based services in general.

Raw trajectory samples obtained from GPS devices are typically of the form of (longitude, latitude, time). If trajectories on network space, we assume that all trajectory sample points have already been map matched onto the vertices of the spatial network using some map-matching algorithm (e.g., [103]) and that between two adjacent sample points p_a and p_b, the object movement always follows the shortest path connecting p_a and p_b. Specifically, a spatial-textual trajectory is defined as follows [104–107].

---

[1]https://www.bing.com/maps/

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

[3]https://www.mapquest.com

[4]https://www.bikely.com/

[5]https://www.gps-waypoints.net

[6]https://www.sharemyroutes.com/

[7]https://research.microsoft.com/en-us/projects/geolife/

[8]https://www.twitter.com/

[9]https://www.Facebook.com/

[10]https://www.Foursquare.com/

Definition (**Trajectory**) A trajectory τ of a moving object is a finite sequence $\langle v_1, v_2, ..., v_n \rangle$, where $v_i = (p_i, t_i)$, $i \in [1, n]$, with p_i being a sample point and t_i being a timestamp (optional).

5.1.2 Trajectory search over evolving road networks

This category of studies perform trajectory search or route planning on evolving road networks. In particular, the weight of edges is time-dependent or changing over time. Hu et al. [108] study the problem of capturing the dependence among eco-weights of adjacent edges on road networks. A number of histogram aggregation algorithms are developed for estimating GHG emissions of routes. Dai et al. [109] focus on estimating past cost distribution by using a collection of past trajectories. Specifically, given a departure time and a query path, the problem is to select an optimal set of weights with associated paths that cover the query path. Here, the weights are required to be the most accurate joint cost distribution estimation for the query path. Yang et al. [110] propose a Path-Centric paradigm that aims at deriving the expected cost of a given path. This paradigm is able to avoid route split by considering the following two sub-problems: (1) calculating the travel cost distribution of a path and (2) discovering a source–destination pair.

5.1.3 Trajectory similarity search and join

Trajectory similarity join is a fundamental problem in the area of trajectory data management. Specifically, given a collection of trajectory, trajectory similarity join is to find all trajectory pairs (e.g., $\langle \tau_i, \tau_j \rangle$) where the similarity between τ_i and τ_j is above a pre-defined similarity threshold θ. Note that the computation of trajectory similarity can be varied. Existing studies apply various trajectory similarity metrics that take spatial proximity and temporal proximity into consideration. To measure the similarity between two trajectories, some studies apply a time interval constraint to constrain the temporal proximity of two trajectories (i.e., [111–114]). According to the attribute of time window, these studies can be classified into two categories: (1) threshold-based constraint [112, 114]; (2) similarity-based constraint [111, 113]. The threshold-based constraint only considers trajectory point pairs whose temporal proximity (time interval) does not exceed the specified threshold. In contrast, the similarity-based constraint use a similarity score as constraint by considering both spatial and temporal aggregate proximities. A host of studies exist that focus on developing effective similarity metrics to measure the similarity between two trajectories. For example, Li et al. [115] propose a deep representation learning based method to compute the similarity between trajectories. Yao et al. [116] further improve the efficiency of neural network based trajectory similarity computation. Xie et al. [117] propose a distributed trajectory search framework to handle trajectory similarity search over massive trajectories. Their framework is implemented in Spark and it supports both the Hausdorff distance the Fréchet distance.

Ta et al. [118] devise a bi-directional mapping similarity (BDS) metric. Given two trajectories τ_1 and τ_2, the idea of BDS is to match each sample point of τ_1 with its closest point on τ_2, and vice versa. Their study claims that it is expensive to enumerate every two trajectories and compute their similarity. To address their challenge, they propose a signature-based trajectory similarity join framework. In particular, at the beginning it generates signatures for each trajectory. After that, a pruning strategy is proposed to filter out unqualified trajectory pairs. Specifically, if two trajectories do not share common signatures, they cannot be similar. Shang et al. [119, 120] develop a divide-and-conquer search framework to process

trajectory similarity join. The framework consists of two phases. In the first phase, each trajectory is selected as a center and all other trajectories similar to the center trajectory is retrieved. techniques to achieve efficiency. In the second phase, the similar trajectories of each center trajectory are merged together. This framework is parallelizable because the searches from each trajectory are independent of each other and can be performed in parallel.

The problem of Trajectory-to-Location join (TL-Join) [121] is investigated recently. TL-Join focuses on the matching between trajectories and locations. Specifically, the inputs are a trajectory set T, a location set L, and a threshold θ. the TL-Join finds all trajectory-location pairs from T and L with similarity above θ. The similarity metric considers both spatial proximity and temporal proximity. They develop a parallel framework to perform the joining operations in parallel.

5.2 Spatial keyword trajectory search

5.2.1 Term-tagged trajectory data

Basically, a term-tagged trajectory is a trajectory augmented with textual description. Two types of term-tagged trajectory data exist: *individually term-tagged trajectory data* and globally term-tagged trajectory data.

Individually term-tagged trajectory (e.g., Han et al. [122]) augments each sample point with keywords, which is defined as follows:

Definition (**Individually term-tagged trajectory**) An individually term-tagged trajectory τ of a moving object is a finite sequence $\langle v_1, v_2, ..., v_n \rangle$, where $v_i = (p_i, t_i, k_i)$, $i \in [1, n]$, with p_i being a sample point, t_i being a timestamp (optional), and k_i being a set of keywords.

In contrast, globally term-tagged trajectory (e.g., Shang et al. [123]) augments the entire trajectory with keywords, which is defined as follows:

Definition (**Globally term-tagged trajectory**) An individually term-tagged trajectory τ of a moving object is a finite sequence sample points $\langle p_1, p_2, ..., p_n \rangle$ and a set K_τ of keywords that describe the textual attributes of τ, e.g., highway, tollway, off-road, travel style, and transport.

5.2.2 Problem formulation

We introduce the existing studies of spatial-keyword queries in trajectory databases. Comparisons of existing spatial-keyword trajectory queries are detailed in Table 1.

User Oriented Trajectory Search (UOTS) is proposed by Shang et al. [123]. Given a set T of spatial-textual trajectories and a query q that includes a set of query locations and a set of keywords, the UOTS query finds the trajectory $\tau \in T$ with the minimum spatial-textual distance $ST_{dist}(q, \tau)$, i.e., $\forall \tau' \in T \setminus \{\tau\}$ $(ST_{dist}(q, \tau) \leq ST_{dist}(q, \tau'))$. The UOTS query is conducted on network space and the spatial similarity is defined by summarizing the network distances between query locations and a trajectory. In the textual domain, approximation keyword matching is adopted and keyword sets are transferred to high dimensional vectors by using the TF-IDF encoding method. Hence, the textual similarity is defined by the distance between two vectors.

Table 1 Comparisons of existing geo-textual trajectory queries

Query	Type	Space	Textual	Temporal
UOTS [123]	NN	Network	Approximation	No
TkSK [104]	kNN	Euclidean	Boolean exact	No
ATSQ [105]	kNN	Euclidean	Boolean exact	No
TkSK [107]	kNN	Euclidean	Approximation	No
AKQST [106]	kNN	Euclidean	Approximation	No
SKRT [122]	Range	Euclidean	Boolean exact	Yes

Example Given a set of sightseeing places and a set of keywords that describe user's pref-erence (tollway: 0%, highway: > 70%, off-road: < 10%, travel style: grouped, transport: by private vehicle), find the trajectory that is spatially close to the sightseeing places and is textually similar to the user's preference.

Top-k Spatial Keyword query (TkSK) is first proposed by Cong et al. [104], and then is extended by Zheng et al. [105–107]. Given a set T of spatial-textual trajectories and a query q that includes a set of query locations and a set of keywords, a top-k spatial key-word trajectory query (TkSK) returns k trajectories from T that have the smallest minimum match distances with respect to q, each associated with the start and end place indexes that yield the minimum match distance. The TkSK query is conducted in Euclidean space and the minimum match distances takes (1) the Euclidean distance between query locations and a trajectory and (2) the travel distance of trajectory (the sum of the matched points) into account. In the textual domain, boolean exact keyword matching or an approximation matching (Edit distance) is adopted.

Example Find the trajectory with the smallest minimum match distances that fully covers the sightseeing places of cave, waterfall, meadow, panda, and kiosk.

Spatial Keyword Range search on Trajectories (SKRT) is proposed by Han et al. [122]. Given a set T of spatial-textual trajectories and a query q that includes a spatial range R, a timespan $[t_s, t_e]$, and a set K of keywords, the SKRT query finds all (sub)trajectories locate within region during query timespan $[t_s, t_e]$, and collectively contain query keywords. The TkSK query is conducted in Euclidean space and boolean exact keyword matching is adopted in the textual domain.

Example Find trajectories in a nearby region during 3:00 pm to 5:00 pm with activities of enjoying wonderful local flower and pizza.

5.3 Methodology

We present the methodologies of the existing spatial-keyword trajectory queries. The comparisons of them is detailed in Table 2.

Collaborative search Shang et al. [123] propose a collaborative algorithm to process the UOTS query efficiently. In the spatial domain, network expansion is adopted to explore spatial networks and to find trajectories spatially close to the query locations. In the textual domain, keyword sets are transferred to high dimensional vectors by using TF-IDF, and

Table 2 Methodologies of geo-textual trajectory queries

Query	Spatial	Textual	Combination
UOTS [123]	Expansion	iDistance	seperate
TkSK [104]	B+-Tree	inverted file	spatial-first
ATSQ [105]	Grid	inverted file	spatial-first
TkSK [107]	R-Tree	inverted file	spatial-first
AKQST [106]	Grid	signature	spatial-first
SKRT [122]	B+-Tree	inverted file	spatial-first

these vectors are indexed by iDistance in high dimensional space. An upper and a lower bound of the spatial-textual distance is defined to prune the search space, and a heuristic search strategy (best-first) is proposed to schedule multiple query sources (query locations and a vector) in both domains effectively.

Hybrid index and optimizations The hybrid index structure for spatial-textual trajectory queries is first proposed by Cong et al. [104] and then Zheng et al. [105–107] study its variants for different applications. The typical procedure of spatial-textual trajectory query processing includes two steps. First, a hybrid spatial-textual index structure is defined, such as B^{ck}-tree [104] (B+-tree plus inverted file), Grid index for Activity Trajectories [105] (GAT, grid index plus inverted file), ITB Tree [107] (R-tree plus inverted file), Grid Keyword index [106] (GiKi, grid index plus keyword signature), and IOC tree [122] (B+-tree plus inverted file). Second, a spatial-first search follows the branch-and-bound strategy to prune the search space and to retrieve the query results.

5.4 Route search and planning

Traditional route search and travel planning Travel planning (a.k.a. route planning) has been playing an indispensable role in transportation. It is a widely studied problem in the areas of data management and GIS.

The problem of travel planning takes trajectory data as input. It outputs an optimal route or a set of routes that are optimized for each traveler. The problem can be classified as trajectory-to-point search and trajectory-to-trajectory search. In the trajectory-to-point search, the problem aims at discovering POIs that are spatially close to a query route according to a similarity metric. The most common problem in this category is the in-route nearest neighbor (IRNN) query [124]. The IRNN query is to find a route close to a fixed route defined by the traveler. Based on the IRNN query, many extension queries are proposed. For example, the path nearest neighbor (PNN) query [125–127] is an extension of the IRNN query, which continuously maintains path nearest neighbors as the traveler is traveling along a pre-defined route. Recently, Shang et al. [128] define a cluster-based travel planning query, the path nearby cluster query, which extends the PNN query by discovering the point clusters that are close to a pre-defined query path.

Different from trajectory-to-point search, trajectory-to-trajectory search is to find trajectories that are similar to a query trajectory. Note that the similarity is measured by taking both geometrical similarity and spatial proximity into consideration. Chen et al. [129] and Shang et al. [102] are the first to study the problem of trajectory similarity search in Euclidean space and road netowrks, respectively. Recently, Yuan et al. [130] propose a distributed trajectory similarity search and join framework on spatial networks.

Shang et al. [131] focus on the problem of finding an optimal travel (ridesharing) arrangement for a set of users located at difference places but have a uniform destination, which is named as collective travel planning (CTP). In particular, given the locations of a set of travelers Q, a set of meeting points S, a uniformed destination d, and an integer threshold k, the CTP problem is to identify a subset A of S with at most k elements that when used as meeting points results in the minimum global travel cost. The CTP query is Max SNP-hard. To find the optimal travel arrangement, they present an exact algorithm by enumerating all possible subsets and an approximation algorithm with polynomial time complexity and and a bounded approximation ratio of 5.

Keyword-aware route search The problem of keyword-aware route search is to return an optimal route satisfying user's keyword-based and threshold-based requirements. A keyword-aware route query $q = \langle s, d, o, c, f \rangle$ [132–137] is basically defined on the basis of a set of geo-tagged items on a spatial network G. In particular, s is the source (starting location) of a route in G, d is the destination (ending location), ø denotes an ordering graph, which is regarded as a directed acyclic graph where each vertex is mapped to a keyword and each edge represents that the keyword at the starting point of the edge is required to be visited before the keyword at the ending point of edge. Note that the keyword matching scheme may be exact matching or approximate matching. Specifically, exact matching denotes that a keyword in an item must be exactly the same as the keyword in a query. In contrast, approximate matching introduces a similarity threshold where the similarity is measured by some string similarity metrics (e.g., edit distance) to determine whether a term in an item matches a term in a query. Δ is a budget limit (e.g., travel distance threshold), which is optional, and f is a function that calculates the objective score of a route (e.g., route popularity score). The query returns a path R in G starting at s and ending at d, such that R optimizes $f(R)$ under the constraints that R satisfies the budget limit Δ and passes through locations following the sequence in Ψ.

Location-based route recommendations Given a set of locations (e.g., POIs, taxi locations), the location-based route recommendation problem aims to derive a new route based on location information, user preference information, and/or other moving patterns. Ge et al. [138] and Ye et al. [139, 140] focus on the mobile sequential recommendation problem, which returns a recommendation result (i.e., a route) by minimizing the potential travel distance to a taxi driver's next potential passenger. The objective of the above problem is to discover an optimal route to a user. Recently, the problem of multiple mobile sequential recommendation is defined and studied. Specifically, Ye et al. [141] is the first to define the problem of multiple mobile sequential recommendation that derives a set of routes by optimizing the travel cost for a group of taxis at different locations.

Apart from mobile sequential recommendation, travel itinerary recommendation is investigated by recent literature (e.g., [142, 143]). In particular, given a set of user-specified POIs and a set of constraints, the problem of travel itinerary recommendation is to generate an itinerary that contains a subset of user-specified POIs, which is named as *refined POI set* with a specific source POI and a specific tatget POI. Here, the output travel route containing the refined POI set is required to be completed within a pre-specified time period. Moreover, Yang et al. [144] focus on the problem of recommending the shortest route to users based on existing trajectories by considering costs of different categories. Additionally, the problem of combined route search by locations is proposed [145]. Given a collection of trajectories, a query POI set Q, and a threshold θ, it returns a combination of trajectories whose similarity to Q is no less than θ the number of combinations should be minimized.

Yawalkar et al. [146] study the problem of finding an optimal route s.t. (1) the length of the route is minimized; (2) the route is expected to cover as many of the query locations as possible. A goal-directed search algorithm is developed for balancing the trade-off between route length and the location coverage. Li et al. [147] propose a differentially private frequent time-constrained sequential pattern mining mechanism, which can be used for route recommendation.

6 Conclusions and future directions

In this paper, we conduct an all-around survey on existing studies of location-based keyword search. We classify existing works of location-based keyword search based on the types of their input data, output results, and methodologies. For each category, we summarize their common features in terms of input data, output result, indexing scheme, and search algorithms. Detailed description regarding each study of spatial keyword search is elaborated.

Though massive studies regarding spatial keyword search exist, we still have a great number of open problems. Here, we summarize two directions that are well worthy of being investigated.

Management of high-dimensional geo-textual data Existing studies use term-based similarity metrics (e.g., language model, BM2.5, cosine similarity) to measure the textual similarity between two items. However, there metrics are computationally expensive, especially when the number of item pair candidates are very large. As a result, we need to enhance computation of on high-dimensional geo-textual data management and query processing in order to provide web users and social-media service providers with more efficient and more scalable mechanism in storing and querying over a much broader range of big multi-modal data. Particular challenge in this field is the exponentially growing time complexity and space complexity in data management and query processing, respectively, when we increase the number of dimensions of source data. , It will be of great interest to extend prior hybrid indexing structure, search algorithms, and benchmark systems to handle high-dimensional streaming data (i.e., with at least 1,000 dimensions) in order to develop more general, efficient, and scalable indexing and query processing mechanisms.

Mining of latent relationship from big geo-textual data streams It is an open problem to effectively and efficiently support real-time mining of relationships between different topics or events in data streams. For example, instead of receiving individual tweets, events, or topics from a stream, users may want to be notified in real time of casual relationships among tweets, events, or topics. Furthermore, high velocity data streams from social media call for distributed solutions. Additionally, data streams can be integrated with static geo-textual objects (e.g., points of interest). By bridging dynamic data streams and static geo-textual data, exciting opportunities for data analytics emerge.

References

1. Chen L, Shang S (2019) Region-based message exploration over spatio-temporal data streams. In: AAAI

2. Yang C, Chen L, Shang S, Zhu F, Li L, Shao L (2019) Toward efficient navigation of massive-scale geo-textual streams. In: IJCAI

3. Cao X, Chen L, Cong G, Jensen CS, Qiang Q, Skovsgaard A, Dingming W, Yiu ML (2012) Spatial keyword querying. In: ER, pp 16–29

4. Bao J, Zheng Y, Wilkie D, Mokbel MF (2015) Recommendations in location-based social networks: a survey. GeoInformatica 19(3):525–565

5. Kong X, Li M, Ma K, Tian K, Wang M, Ning Z, Xia F (2018) Big trajectory data: a survey of applications and services. IEEE Access 6:58295–58306

6. Feng Z, Zhu Y (2016) A survey on trajectory data mining: techniques and applications. IEEE Access 4:2056–2067

7. Mahmood AR, Punni S, Aref WG (2019) Spatio-temporal access methods: a survey (2010 - 2017). GeoInformatica 23(1):1–36

8. Mahmood AR, Aref WG (2019) Scalable processing of spatial-keyword queries. Synth Lect Data Manag 11(1):1–116

9. Cong G, Jensen CS (2016) Querying geo-textual data: spatial keyword queries and beyond. In: SIGMOD, pp 2207–2212

10. Cong G, Feng K, Zhao K (2016) Querying and mining geo-textual data for exploration: challenges and opportunities. In: ICDE Workshops, pp 165–168

11. Chen L, Cong G, Jensen CS, Wu D (2013) Spatial keyword query processing: an experimental evaluation. PVLDB 6(3):217–228

12. Liu Y, Pham T-A, Cong G, Yuan Q (2017) An experimental evaluation of point-of-interest recommendation in location-based social networks. PVLDB 10(10):1010–1021

13. Vaid S, Jones CB, Joho H, Sanderson M (2005) Spatio-textual indexing for geographical search on the web. In: SSTD, pp 218–235

14. Zhou Y, Xie X, Wang C, Gong Y, Ma W-Y (2005) Hybrid index structures for location-based web search. In: CIKM, pp 155–162

15. Chen Y-Y, Suel T, Markowetz A (2006) Efficient query processing in geographic web search engines. In: SIGMOD, pp 277–288

16. Hariharan R, Hore B, Li C, Mehrotra S (2007) Processing spatial-keyword (sk) queries in geographic information retrieval (gir) systems. In: SSDBM, p 16

17. Khodaei A, Shahabi C, Li C (2010) Hybrid indexing and seamless ranking of spatial and textual features of web documents. In: DEXA (1), pp 450–466

18. Christoforaki M, He J, Dimopoulos C, Markowetz A, Suel T (2011) Text vs. space: efficient geo-search query processing. In: CIKM, pp 423–432

19. Zhang D, Tan K-L, Tung AKH (2013) Scalable top-k spatial keyword search. In: EDBT, pp 359–370

20. Felipe ID, Hristidis V, Rishe N (2008) Keyword search on spatial databases. In: ICDE, pp 656–665

21. Cary A, Wolfson O, Rishe N (2010) Efficient and scalable method for processing top-k spatial boolean queries. In: SSDBM, pp 87–95

22. Dingming W, Yiu ML, Cong G, Jensen CS (2012) Joint top-k spatial keyword query processing. IEEE Trans Knowl Data Eng 24(10):1889–1903

23. Tao Y, Sheng C (2014) Fast nearest neighbor search with keywords. IEEE Trans Knowl Data Eng 26(4):878–888

24. Zhang C, Zhang Y, Zhang W, Lin X (2013) Inverted linear quadtree: efficient top k spatial keyword search. In: ICDE, pp 901–912

25. Cong G, Jensen CS, Dingming W (2009) Efficient retrieval of the top-k most relevant spatial web objects. In: PVLDB, pp 337–348

26. Li Z, Lee KCK, Zheng B, Lee W-C, Lee DL, Wang X (2011) Ir-tree: an efficient index for geographic document search, vol 23

27. Rocha-Junior JB, Gkorgkas O, Jonassen S, Nørvåg K (2011) Efficient processing of top-k spatial keyword queries. In: SSTD, pp 205–222

28. Fang H, Zhao P, Sheng VS, Li Z, Jiajie X, Jian W, Cui Z (2015) Ranked reverse boolean spatial keyword nearest neighbors search. In: WISE, pp 92–107

29. Jiaheng L, Ying L, Cong G (2011) Reverse spatial and textual k nearest neighbor search. In: SIGMOD, pp 349–360

30. Ying L, Jiaheng L, Cong G, Wei W, Shahabi C (2014) Efficient algorithms and cost models for reverse spatial-keyword k-nearest neighbor search. ACM Trans Database Syst 39(2):13

31. Ying L, Cong G, Jiaheng L, Shahabi C (2015) Efficient algorithms for answering reverse spatial-keyword nearest neighbor queries. In: SIGSPATIAL, pp 82:1–82:4

32. Gao Y, Qin X, Zheng B, Chen G (2015) Efficient reverse top-k boolean spatial keyword queries on road networks. IEEE Trans Knowl Data Eng 27(5):1205–1218

33. Changyin L, Li J, Li G, Wei W, Li Y, Li J (2016) Efficient reverse spatial and textual k nearest neighbor queries on road networks. Knowl-Based Syst 93:121–134
34. Shang S, Bo Y, Ke D, Xie K, Zhou X (2011) Finding the most accessible locations: reverse path nearest neighbor query in road networks. In: SIGSPATIAL, pp 181–190
35. Cao X, Cong G, Guo T, Jensen CS, Ooi BC (2015) Efficient processing of spatial group keyword queries. ACM Trans Database Syst 40(2):13
36. Cao X, Cong G, Jensen CS, Ooi BC (2011) Collective spatial keyword querying. In: SIGMOD, pp 373–384
37. Long C, Wong RC-W, Wang K, Fu AW-C (2013) Collective spatial keyword queries: a distance owner-driven approach. In: SIGMOD, pp 689–700
38. Zhang D, Ooi BC, Tung AKH (2010) Locating mapped resources in web 2.0. In: ICDE, pp 521–532
39. Guo T, Cao X, Cong G (2015) Efficient algorithms for answering the m-closest keywords query. In: SIGMOD, pp 405–418
40. Zhang D, Chee YM, Mondal A, Tung AKH, Kitsuregawa M (2009) Keyword search in spatial databases: towards searching by document. In: ICDE, pp 688–699
41. Mehta P, Skoutas D, Sacharidis D, Voisard A (2016) Coverage and diversity aware top-k query for spatio-temporal posts. In: SIGSPATIAL, pp 37:1–37:10
42. Drosou M, Pitoura E (2012) Dynamic diversification of continuous data. In: EDBT, pp 216–227
43. Minack E, Siberski W, Nejdl W (2011) Incremental diversification for very large sets: a streaming-based approach. In: SIGIR, pp 585–594
44. Liang S, Yilmaz E, Shen H, de Rijke M, Bruce W (2017) Croft Search result diversification in short text streams. ACM Trans Inf Syst 36(1):8,1–8,35
45. Cheng S, Arvanitis A, Chrobak M, Hristidis V (2014) Multi-query diversification in microblogging posts. In: EDBT, pp 133–144
46. Chen L, Cong G (2015) Diversity-aware top-k publish/subscribe for text stream. In: SIGMOD, pp 347–362
47. Zhang C, Zhang Y, Zhang W, Lin X, Cheema MA, Wang X (2014) Diversified spatial keyword search on road networks. In: EDBT, pp 367–378
48. Derczynski L, Yang B, Jensen CS (2013) Towards context-aware search and analysis on social media data. In: EDBT, pp 137–142
49. Qiang Q, Liu S, Yang B, Jensen CS (2014) Efficient top-k spatial locality search for co-located spatial web objects. In: MDM, pp 269–278
50. Sun J, Jiajie X, Zheng K, Liu C (2017) Interactive spatial keyword querying with semantics. In: CIKM, pp 1727–1736
51. Qian Z, Jiajie X, Zheng K, Zhao P, Zhou X (2018) Semantic-aware top-k spatial keyword queries. World Wide Web 21(3):573–594
52. Qiang Q, Liu S, Yang B, Jensen CS (2014) Integrating non-spatial preferences into spatial location queries. In: SSDBM, pp 8:1–8:12
53. Han J, Zheng K, Sun A, Shang S, Wen J-R (2016) Discovering neighborhood pattern queries by sample answers in knowledge base. In: ICDE, pp 1014–1025
54. Han J, Wen J-R, Pei J (2014) Within-network classification using radius-constrained neighborhood patterns. In: CIKM, pp 1539–1548
55. Han J, Wen J-R (2013) Mining frequent neighborhood patterns in a large labeled graph. In: CIKM, pp 259–268
56. Yang J, Zhang Y, Huiqi H, Xing C (2019) A hierarchical index structure for region-aware spatial keyword search with edit distance constraint. In: DASFAA, pp 591–608
57. Yang J, Zhang Y, Zhou X, Wang J, Huiqi H, Xing C (2019) A hierarchical framework for top-k location-aware error-tolerant keyword search. In: ICDE, pp 986–997
58. Zhao S, King In, Lyu MR (2016) A survey of point-of-interest recommendation in location-based social networks. arXiv:abs/1607.00647
59. Wang H, Terrovitis M, Mamoulis N (2013) Location recommendation in location-based social networks using user check-in data. In: SIGSPATIAL, pp 364–373
60. Ziyu L, Wang H, Mamoulis N, Wenting T, Cheung DW (2017) Personalized location recommendation by aggregating multiple recommenders in diversity. GeoInformatica 21(3):459–484
61. Ye M, Yin P, Lee W-C, Lee DL (2011) Exploiting geographical influence for collaborative point-of-interest recommendation. In: SIGIR, pp 325–334
62. Zhang J-D, Chow C-Y (2013) igslr: personalized geo-social location recommendation: a kernel density estimation approach. In: *SIGSPATIAL*, pp 324–333
63. Zhang J-D, Chow C-Y, Li Y (2014) LORE: exploiting sequential influence for location recommendations. In: SIGSPATIAL, pp 103–112

64. Cheng C, Yang H, King I, Lyu MR (2012) Fused matrix factorization with geographical and social influence in location-based social networks. In: AAAI
65. Liu B, Xiong H, Papadimitriou S, Yanjie F, Yao Z (2015) A general geographical probabilistic factor model for point of interest recommendation. IEEE Trans Knowl Data Eng 27(5):1167–1179
66. Liu Y, Wei W, Sun A, Miao C (2014) Exploiting geographical neighborhood characteristics for location recommendation. In: CIKM, pp 739–748
67. Lian D, Zhao C, Xie X, Sun G, Chen E, Rui Y (2014) Geomf: joint geographical modeling and matrix factorization for point-of-interest recommendation. In: KDD, pp 831–840
68. Li X, Cong G, Li X, Pham T-AN, Krishnaswamy S (2015) Rank-geofm: a ranking based geographical factorization method for point of interest recommendation. In: SIGIR, pp 433–442
69. Li H, Ge Y, Hong R, Zhu H (2016) Point-of-interest recommendations: learning potential check-ins from friends. In: KDD, pp 975–984
70. Gao H, Tang J, Xia H, Liu H (2013) Exploring temporal effects for location recommendation on location-based social networks. In: RecSys, pp 93–100
71. Ma C, Zhang Y, Wang Q, Liu X (2018) Point-of-interest recommendation: exploiting self-attentive autoencoders with neighbor-aware influence. In: CIKM, pp 697–706
72. Sen S, Yan H, Cheng X, Tang P, Peng X, Jianliang X (2015) Authentication of top-k spatial keyword queries in outsourced databases. In: DASFAA, pp 567–588
73. Dingming W, Choi B, Jianliang X, Jensen CS (2015) Authentication of moving top-k spatial keyword queries, vol 27
74. Obiri IA, Wang Y, Nuhoho RE, Owiyo E (2018) Authentication of multiple-user spatial keywords queries. In: DSC, pp 506–513
75. Yue X, Xi M, Chen B, Gao M, He Y, Xu J A revocable group signatures scheme to provide privacy-preserving authentications. In: Mobile networks and applications, pp 1–30, online first
76. Sen S, Teng Y, Cheng X, Wang Y, Li G (2015) Privacy-preserving top-k spatial keyword queries over outsourced database. In: DASFAA, pp 589–608
77. Sen S, Teng Y, Cheng X, Ke X, Li G, Chen J (2018) Privacy-preserving top-k spatial keyword queries in untrusted cloud environments. IEEE Trans Serv Comput 11(5):796–809
78. Cui N, Li J, Yang X, Wang B, Reynolds M, Xiang Y (2019) When geo-text meets security: privacy-preserving boolean spatial keyword queries. In: ICDE, pp 1046–1057
79. Li G, Wang Y, Wang T, Feng J (2013) Location-aware publish/subscribe. In: KDD, pp 802–810
80. Chen L, Cong G, Cao X (2013) An efficient query indexing mechanism for filtering geo-textual data. In: SIGMOD, pp 749–760
81. Magdy A, Mokbel MF, Elnikety S, Nath S, He Y (2014) Mercury: a memory-constrained spatio-temporal real-time search on microblogs. In: ICDE, pp 172–183
82. Minghe Y, Li G, Feng J (2015) A cost-based method for location-aware publish/subscribe services. In: CIKM, pp 693–702
83. Minghe Y, Li G, Wang T, Feng J, Gong Z (2015) Efficient filtering algorithms for location-aware publish/subscribe. IEEE Trans Knowl Data Eng 27(4):950–963
84. Chen L, Cong G, Cao X, Tan K-L (2015) Temporal spatial-keyword top-k publish/subscribe. In: ICDE, pp 255–266
85. Wang X, Zhang Y, Zhang W, Lin X, Wang W (2015) Ap-tree: efficiently support continuous spatial-keyword queries over stream. In: ICDE, pp 1107–1118
86. Huiqi H, Liu Y, Li G, Feng J, Tan K-L (2015) A location-aware publish/subscribe framework for parameterized spatio-textual subscriptions. In: ICDE, pp 711–722
87. Wang X, Zhang Y, Zhang W, Lin X, Wang W (2015) Ap-tree: efficiently support location-aware publish/subscribe. VLDB J 24(6):823–848
88. Wang X, Zhang Y, Zhang W, Lin X, Huang Z (2016) SKYPE: top-k spatial-keyword publish/subscribe over sliding window. PVLDB 9(7):588–599
89. Wang X, Zhang W, Zhang Y, Lin X, Huang Z (2017) Top-k spatial-keyword publish/subscribe over sliding window. VLDB J 26(3):301–326
90. Chen Z, Cong G, Zhang Z, Tom Z, Fu J, Chen L (2017) Distributed publish/subscribe query processing on the spatio-textual data stream. In: ICDE, pp 1095–1106
91. Xiong X, Mokbel MF, Aref WG (2017) Continuous queries in spatio-temporal databases. In: Encyclopedia of GIS, pp 340–345
92. Chen L, Shang S (2018) Approximate spatio-temporal top-k publish/subscribe. World Wide Web, 1–23
93. Skovsgaard A, Sidlauskas D, Jensen CS (2014) Scalable top-k spatio-temporal term querying. In: ICDE, pp 148–159
94. Mahmood AR, Aly AM, Aref WG (2018) FAST: frequency-aware indexing for spatio-textual data streams. In: ICDE, pp 305–316

95. Mahmood AR, Aref WG, Aly AM (2017) FAST: frequency-aware spatio-textual indexing for in-memory continuous filter query processing. arXiv:abs/1709.02529
96. Ying X, Chen L, Yao B, Shang S, Zhu S, Zheng K, Li F (2017) Location-based top-k term querying over sliding window. In: WISE, pp 299–314
97. Chen L, Shang S, Yao B, Zheng K (2018) Spatio-temporal top-k term search over sliding window. World Wide Web, 1–18
98. Chen L, Shang S, Zhang Z, Cao X, Jensen CS, Kalnis P (2018) Location-aware top-k term publish/subscribe. In: ICDE, pp 749–760
99. Guo L, Zhang D, Li G, Tan K-L, Bao Z (2015) Location-aware pub/sub system: when continuous moving queries meet dynamic event streams. In: SIGMOD, pp 843–857
100. Zhao K, Chen L, Cong G (2016) Topic exploration in spatio-temporal document collections. In: SIGMOD, pp 985–998
101. Chen L, Shang S, Zheng K, Kalnis P (2019) Cluster-based subscription matching for geo-textual data streams. In: ICDE, pp 890–901
102. Shang S, Ding R, Zheng K, Jensen CS, Kalnis P, Zhou X (2014) Personalized trajectory matching in spatial networks. VLDB J 23(3):449–468
103. Brakatsoulas S, Pfoser D, Salas R, Wenk C (2005) On map-matching vehicle tracking data. In: VLDB, pp 853–864
104. Cong G, Lu H, Ooi BC, Zhang D, Zhang M (2012) Efficient spatial keyword search in trajectory databases. arXiv:1205.2880, 1–12
105. Zheng K, Shang S, Yuan NJ, Yi Y (2013) Towards efficient search for activity trajectories. In: ICDE, pp 230–241
106. Zheng B, Yuan NJ, Zheng K, Xie X, Sadiq SW, Zhou X (2015) Approximate keyword search in semantic trajectory database. In: ICDE, pp 975–986
107. Zheng K, Zheng B, Xu J, Liu G, Liu A, Li Z (2017) Popularity-aware spatial keyword search on activity trajectories. World Wide Web, online first, 1–25
108. Jilin H, Yang B, Jensen CS, Ma Y (2017) Enabling time-dependent uncertain eco-weights for road networks. GeoInformatica 21(1):57–88
109. Dai J, Yang B, Guo C, Jensen CS, Hu J (2016) Path cost distribution estimation using trajectory data. PVLDB 10(3):85–96
110. Yang B, Dai J, Guo C, Jensen CS, Hu J (2018) PACE: a path-centric paradigm for stochastic path finding. VLDB J 27(2):153–178
111. Bakalov P, Hadjieleftheriou M, Keogh EJ, Tsotras VJ (2005) Efficient trajectory joins using symbolic representations. In: MDM, pp 86–93
112. Bakalov P, Tsotras VJ (2006) Continuous spatiotemporal trajectory joins. In: GSN, pp 109–128
113. Chen Y, Patel JM (2009) Design and evaluation of trajectory join algorithms. In: SIGSPATIAL, pp 266–275
114. Ding H, Trajcevski G, Scheuermann P (2008) Efficient similarity join of large sets of moving object trajectories. In: TIME, pp 79–87
115. Li X, Zhao K, Cong G, Jensen CS, Wei W (2018) Deep representation learning for trajectory similarity computation. In: ICDE, pp 617–628
116. Di Y, Cong G, Zhang C, Bi J (2019) Computing trajectory similarity in linear time: A generic seed-guided neural metric learning approach. In: ICDE, pp 1358–1369
117. Xie D, Li F, Phillips JM (2017) Distributed trajectory similarity search. PVLDB 10(11):1478–1489
118. Na T, Li G, Xie Y, Li C, Hao S, Feng J (2017) Signature-based trajectory similarity join. IEEE Trans Knowl Data Eng 29(4):870–883
119. Shang S, Chen L, Wei Z, Jensen CS, Zheng K, Kalnis P (2017) Trajectory similarity join in spatial networks. PVLDB 10(11):1178–1189
120. Shang S, Chen L, Wei Z, Jensen CS, Zheng K, Kalnis P (2018) Parallel trajectory similarity joins in spatial networks. VLDB J 27(3):395–420
121. Shang S, Chen L, Zheng K, Jensen CS, Wei Z, Kalnis P (2019) Parallel trajectory-to-location join. IEEE Trans Knowl Data Eng 31(6):1194–1207
122. Han Y, Wang L, Zhang Y, Zhang W, Lin X (2015) Spatial keyword range search on trajectories. In: DASFAA, pp 223–240
123. Shang S, Ding R, Bo Y, Xie K, Zheng K, Kalnis P (2012) User oriented trajectory search for trip recommendation. In: EDBT, pp 156–167
124. Shekhar S, Yoo JS (2003) Processing in-route nearest neighbor queries: a comparison of alternative approaches. In: GIS, pp 9–16
125. Chen Z, Shen HT, Zhou X, Xu JY (2009) Monitoring path nearest neighbor in road networks. In: SIGMOD, pp 591–602

126. Shang S, Ke D, Xie K (2010) Best point detour query in road networks. In: SIGSPATIAL, pp 71–80
127. Shang S, Bo Y, Ke D, Xie K, Zheng K, Zhou X (2012) PNN query processing on compressed trajectories. GeoInformatica 16(3):467–496
128. Shang S, Zheng K, Jensen CS, Yang B, Kalnis P, Li G, Wen J-R (2015) Discovery of path nearby clusters in spatial networks. IEEE Trans Knowl Data Eng 27(6):1505–1518
129. Chen Z, Shen HT, Zhou X, Zheng Y, Xie X (2010) Searching trajectories by locations: an efficiency study. In: SIGMOD, pp 255–266
130. Yuan H, Li G (2019) Distributed in-memory trajectory similarity search and join on road network. In: 2019 IEEE 35th international conference on data engineering (ICDE), pp 1262–1273
131. Shang S, Chen L, Wei Z, Jensen CS, Wen J-R, Kalnis P (2016) Collective travel planning in spatial networks. IEEE Trans Knowl Data Eng 28(5):1132–1146
132. Li J, Yang YD, Mamoulis N (2013) Optimal route queries with arbitrary order constraints. IEEE Trans Knowl Data Eng 25(5):1097–1110
133. Cao X, Chen L, Cong G, Xiao X (2012) Keyword-aware optimal route search. PVLDB 5(11):1136–1147
134. Cao X, Chen L, Cong G, Guan J, Phan N-T, Xiao X (2013) KORS: keyword-aware optimal route search system. In: ICDE, pp 1340–1343
135. Li Y, Yang W, Dan W, Xie Z (2015) Keyword-aware dominant route search for various user preferences. In: DASFAA, pp 207–222
136. Yao B, Tang M, Li F (2011) Multi-approximate-keyword routing in GIS data. In: SIGSPATIAL, pp 201–210
137. Li M, Chen L, Cong G, Yu G, Ge Y (2016) Efficient processing of location-aware group preference queries. In: CIKM, pp 559–568
138. Ge Y, Xiong H, Tuzhilin A, Xiao K, Gruteser M, Pazzani MJ (2010) An energy-efficient mobile recommender system. In: KDD, pp 899–908
139. Ye Z, Xiao K, Ge Y, Deng Y (2019) Applying simulated annealing and parallel computing to the mobile sequential recommendation. IEEE Trans Knowl Data Eng 31(2):243–256
140. Ye Z, Xiao K, Deng Y (2018) A unified theory of the mobile sequential recommendation problem. In: ICDM, pp 1380–1385
141. Ye Z, Zhang L, Xiao K, Zhou W, Ge Y, Deng Y (2018) Multi-user mobile sequential recommendation: an efficient parallel computing paradigm. In: KDD, pp 2624–2633
142. Lim KH, Chan J, Karunasekera S, Leckie C (2017) Personalized itinerary recommendation with queuing time awareness. In: SIGIR, pp 325–334
143. Taylor K, Lim KH, Chan J (2018) Travel itinerary recommendations with must-see points-of-interest. In: WWW, pp 1198–1205
144. Dai J, Yang B, Guo C, Ding Z (2015) Personalized route recommendation using big trajectory data. In: ICDE, pp 543–554
145. Chen L, Shang S, Jensen CS, Yao B, Zhang Z, Shao L (2019) Effective and efficient reuse of past travel behavior for route recommendation. In: KDD
146. Yawalkar P, Ranu S (2019) Route recommendations on road networks for arbitrary user preference functions. In: ICDE, pp 602–613
147. Li Y, Wang G, Ye Y, Cao X, Yuan L, Lin X (2018) Privts: differentially private frequent time-constrained sequential pattern mining. In: DASFAA, pp 92–111

Lisi Chen obtained his Ph.D. in computer science from Nanyang Technological University, Singpore, in 2016. His research interests include geo-textual data management, spatial keyword query evaluation, and location based social networks. His research aims to build up personalized information dissemination systems of high quality, efficiency, and scalability.

Shuo Shang obtained his Ph.D. in computer science from The University of Queensland, Australia. His research interests include efficient query processing in spatiotemporal databases, spatial trajectory computing, and location based social media. He has served as PC member, session chair, and invited reviewer for many prestigious conferences and journals, including SIGMOD, ICDE, TKDE, The VLDB Journal, ACM TIST, IEEE TITS, ACM TSAS, GeoInformatica, KAIS, WWW Journal, DKE, JCST, and IEICE Transactions.

Chengcheng Yang is a research associate at Inception Institute of Artificial Intelligence, UAE. He obtained his Ph.D. in computer science from the University of Science and Technology of China. His research interests include database storage systems and data compression.

Jing Li is a Research Scientist at the Inception Institute of Artificial Intelligence, UAE. He obtained his PhD degree in Computer Science, at Nanyang Technological University, Singapore. His research aims to build up semantic Web systems to support information needs of web users via deep text understanding, information extraction, machine intelligent question answering, knowledge representation, as well as social media analysis.