Article

Urban Analytics and City Science

EPB: Urban Analytics and City Science
0(0) 1–19

DOI: 10.1177/2399808320913300
journals.sagepub.com/home/epb
SAGE

Multi-objective trajectory optimization in planning for sequential activities across space and through time

Xin Feng
University of California at Santa Barbara, USA

Shaohua Wang
University of California at Santa Barbara, USA; Institute of Geographic Sciences and Natural Resources Research, China Academy of Science, China

Alan T Murray
University of California at Santa Barbara, USA

Yuanpei Cao
Airbnb Incorporation, USA

Song Gao
University of Wisconsin, Madison, USA

Abstract

Human movement and interaction across space and through time is full of economic and social opportunities. Access to information through location-based technologies offers potential for people to make better decisions about social activity participation needs and travel behavior preferences. Identifying an optimal trajectory (route) connecting desired activity locations for multiple attendees with space–time constraints is a challenging endeavor. This spatial organization task is formulated mathematically as a sequential, multi-objective optimization model. A framework consisting of context knowledge, geographic information systems, and spatial optimization is structured to solve this model, allowing for the integration of geographical and social networking considerations. The proposed approach offers a way to balance the tradeoffs of many participants, enabling explicit consideration of travel cost, personal preference, quality rating, etc.

Corresponding author:
Shaohua Wang, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, 11A, Datun Road, Chaoyang District, Beijing 100101, China.
Email: wangshaohua@lreis.ac.cn

in activity planning and decision making. A case study is detailed involving the organization of multiple activities and multiple individuals. The application results highlight the utility and insight of the proposed model and associated solution approaches.

Keywords
Route planning, network analysis, sequential spatial optimization

Introduction

Human movement and interaction in space and time play an important role in daily life. People crave interaction with others on a regular basis. They seek to build connections in order to acquire information, materials, goods, or satisfy emotional needs. Work, school, and other duties require movement through a region within a specific time window, also known as the "coupling constraints" in time geography (Hägerstrand, 1970). These interactions are often spread throughout the day in a sequential manner. Society functions based on such movements and meetups. Further, each individual has unique needs and associated travel behavior. In order to organize multiple activities, people implicitly or explicitly optimize meetup locations, travel costs, and preferences. Greater efficiency and organization can enhance the utility associated with activities at different meetup locations and at various times throughout the day.

Ever increasing location-based technologies enable people to gain access to rich information to support decision making about social activity participation. For example, Google Maps has the ability to check possible routes and landmarks for faster travel times based on trips by car, bike, foot, or public transportation. There have been many studies that develop analytical foundations and insights about human activities, interactions, transportation, path choice, etc. (Kim and Kwan, 2003; Miller, 1991, 2005; Shaw and Yu, 2009). However, research on efficient sequencing of multiple individual movements given space–time constraints is limited. Consider the following situation: a group of friends would like to organize a series of activities in order to celebrate a noteworthy event, such as an anniversary, birthday, or graduation. They expect to have lunch together, go for a walk, and enjoy evening entertainment. The questions raised here are twofold: assuming flexibility, how can we organize these sequential activities in order to minimize travel costs; and, what activities will maximize the overall preferences of participants. The travel of individuals can be viewed as the movement of objects across space and time. Accordingly, what is sought is to optimize meetup locations for these sequential activities to minimize travel cost, enhance preferences, and increase group participation. The determination of the activity locations must be done simultaneously rather than independently, one at a time. Individuals may have different trip chains based on their own constraints but will share several common activities.

Although minimizing travel cost is important, knowing the individual preferences of the attendees is also essential. Budget conscious individuals may prefer to go to a lower cost restaurant while others may prefer more upscale dining options. Some people may prefer a walk over steep and challenging terrain while others may opt for less strenuous hiking. For evening entertainment, some may prefer a bar within walking distance, while others are fine taking a taxi/Uber. The best individual selection of activities may look quite different among participants as they have different preferences, yet acceptable or desirable alternatives no doubt exist when all preferences are considered. Balancing travel cost along with individual preferences is therefore a challenging question worth exploring.

In this research, we propose a multi-objective spatial optimization model to structure and solve the problem of finding sequential activity locations for many people, each with different origins and destinations. Specifically, this problem includes two interrelated parts, selecting meetup locations and specifying the trajectories of travel through them. The contributions of our research are threefold. First, we mathematically formalize and solve this location-routing problem along road networks. Second, a framework consisting of context knowledge, geographic information systems (GIS), and spatial optimization is detailed. The framework allows the integration of geographical contexts (e.g. traffic congestion) and other geospatial information (e.g. points of interest (POI) database, individual preference, statistical rating data, etc.) to be considered in the decision-making process. Third, a multi-objective approach is introduced, accounting for travel cost as well as attendee preference in the trajectory optimization process. Finally, a heuristic algorithm is developed and applied in order to improve computational efficiency of the proposed approach for big data environments.

The paper is structured as follows. The next section reviews relevant literature. Then, a framework for analysis is detailed, incorporating context information, GIS, and spatial optimization components. A case study involving travel within the Phoenix, Arizona, metropolitan region is then introduced. Results are presented associated with selecting sequential activity locations for a group of individuals. The paper ends with discussion and concluding comments.

Background

Research related to issues of activity space coordination has spanned a number of fields, and includes time geography, location modeling, path optimization, social media, etc. In the field of time geography, related advances have focused on evaluating spatial accessibility, feasible opportunity set, and possible activity duration (Kim and Kwan, 2003). Miller (1991, 2005) advocated space–time prism concepts within GIS to describe human activities and interaction with spatiotemporal constraints. Even though space–time prisms capture locations for moving objects, it does not account for inherent flexibility of movements. Kuijpers and Othman (2009) and Kuijpers et al. (2010) were among the first to address this issue by modeling uncertainty in moving objects along a road network. Their concept of “anchor regions” has been widely used in the field of transportation, especially in urban public transportation planning (Song et al., 2017). Research associated with space–time prisms has answered important questions about whether and/or where people are able to meet. However, these studies have limited capacity to identify the best meetup locations. A significant question, therefore, is how to distinguish between meetup locations within accessible areas.

Location modeling has often been used to optimize goods and services by identifying a specific location for a specific activity. Optimizing one’s activity location could be most simply reflected in the Weber problem, where a facility location is sought such that the total weighted travel distance or transportation cost is minimized (see Church and Murray, 2009). The facility and demands in the Weber problem correspond to the activity (or meetup) location and attendees, respectively. The multi-Weber and the p-median are location-allocation problems that have been widely applied in location modeling (Church and Murray, 2009). The multi-Weber problem, proposed initially by Cooper (1963), concerns siting multiple facilities simultaneously to serve regional demands. While accounting for activity location selection, there are major limitations in routing and coordination with such an approach. A key issue is enabling attendees to participate in more than one activity, as the multi-Weber problem does not account for multiple-destination routing in location selection. A new model is therefore needed to site multiple activity locations and optimize attendee trajectories (routes) to go to all activities.

In order to optimize the selection of multiple activity locations along road networks, we need to include shortest path routing. Dijkstra's (1959) algorithm is frequently used to find the shortest path from one origin to one or more destinations in a network. The algorithm implicitly considers all possible routes, though this may be time consuming for large networks. Several algorithms, including convex-hull-based (De Berg et al., 2008), diameter-point-based (Aingworth et al., 1999), R* tree-based spatial indexing (Beckmann et al., 1990), among others, have been used to speed up shortest path computation by reducing the search space of possible routes (Wagner and Willhalm, 2007; Wang et al., 2018). The gateway shortest path model proposed by Lombard and Church (1993) seeks the shortest path that is constrained to travel through one specified node. The multi-gateway shortest path model (Scaparra et al., 2014) has the potential to optimize trajectory for each attendee in the multiple activities location problem. A weakness is that the models focus only on how to route attendees, not sequentially site activity locations. Therefore, coordinating the location of multiple activities along with efficient individual travel routes in a specific order is essential.

Individual preference of potential activity locations makes location selection and travel cost derivation challenging. For example, people who are not in a hurry may be willing to travel further to go to a location that gives them greater satisfaction. Social media could aid related description/measurement of preferences. With the development of location-based social networks (e.g. Foursquare, Yelp, etc.), people are able to share tips and experiences from visits to POIs (e.g. restaurants, coffee shops, bars, etc.) with their friends or other users. Properties of location-based social networks, such as user-item ratings, user check-in data, parking, wheelchair accessibility, price range, kid friendliness, WiFi access, etc. have been considered using various approaches. Accounting for these properties, recommender systems have been widely used and studied (e.g. Wang et al., 2013; Ye et al., 2011). However, what remains a challenge is recommending a series of ordered activity locations, not only considering attendee preferences but also travel costs. The research goal is to integrate preference with travel cost since tradeoffs may exist when accounting for both at the same time. Spatial optimization provides the capacity to balance them, which is precisely the intent of this paper.

Methods

Finding sequential activity locations for multiple moving objects along road networks requires supporting analytics. This paper focuses on spatial analytics, those methods that facilitate systematic exploration of geographic data. To this end, a framework is developed consisting of spatial information, context information, GIS, and spatial optimization (Figure 1). The framework highlights the interaction among decision making, geographic information, and a range of spatial analytics.

Input data (top layer in Figure 1) in the analysis process include detailed spatial information, like streets, road intersections, travel patterns, locations of POIs, origin and destination of each participant, etc. If these data do not exist, then they must be created and/or purchased. Further, the preferences for social interaction are also essential. Satisfaction of need is difficult to measure yet has significant impacts on decision making. By comparing

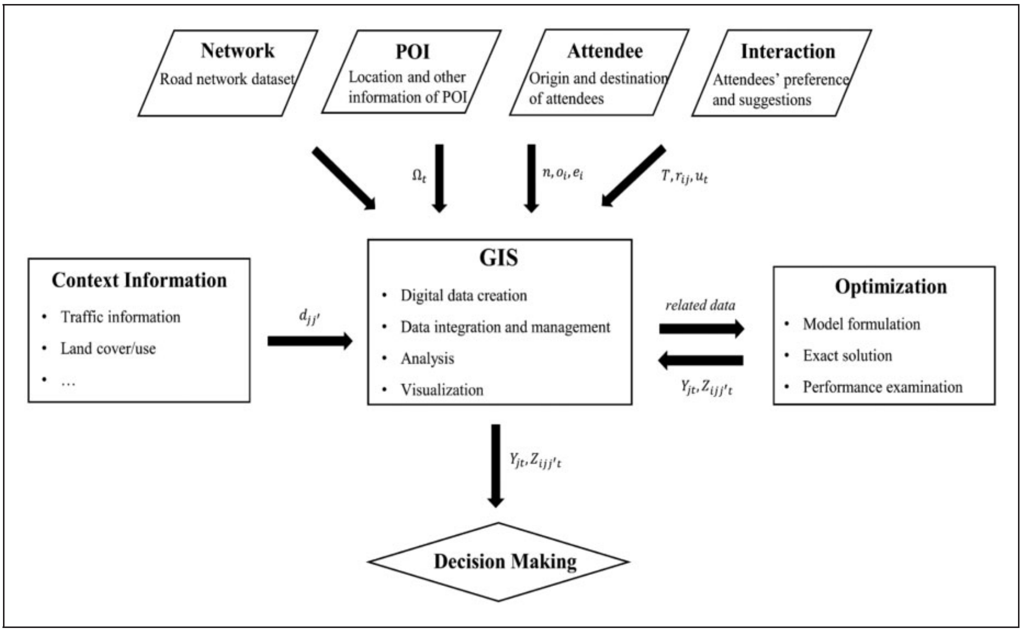

Figure 1. Framework for optimizing sequential activity planning.
GIS: geographic information systems; POI: points of interest.

attendee preference and potential location attributes, satisfaction can be derived for each attendee at any POI location.

An advantage of the proposed framework is integrating geographic context ("Context information" box in Figure 1), providing capacity to support dynamic exploration and planning for activity coordination. Context information, such as traffic congestion, temporal barriers, and other local environmental conditions, has proven to have essential effects on trajectory siting connecting origins and destinations (Buchin et al., 2012; Dodge et al., 2016; Siła-Nowicka et al., 2016). Statistical traffic information on road segments enables context-awareness to be introduced as part of travel cost. For mountainous areas, view shed analysis using digital elevation models may be employed to facilitate travel cost surface estimation. Without context information, road availability, road quality as well as traffic congestion and other road conditions are assumed to be identical for all trips.

Spatial analytics are integral to support this framework. In particular, the combination of GIS and spatial optimization in Figure 1 is useful and meaningful for developing efficient and strategic planning insights. GIS is designed to capture, store, manipulate, analyze, and display all types of spatial/geographical data (Church and Murray, 2009; Clarke, 2011). These capabilities are important components of the framework in Figure 1 for modeling and decision making with respect to activities involving human interaction. GIS supports data creation; in this case, the origin and destination of each participant, the road network, and possible traffic jam locations. GIS facilitates integration and management of different kinds of data as well; these aspects include parcels, roads, transit routes/stops, and different categories of POIs. GIS provides basic spatial analysis capabilities, like measuring the distance between nodes along a network or the density of certain type of POIs in an area. Visualization and display of activity locations and their associated trajectory for all attendees is an unambiguous task that GIS supports.

In order to move beyond basic analysis and visual examination by GIS, other spatial analytical tools are necessary. For example, optimization models in Figure 1 provide the capacity to improve travel efficiency. Wang et al. (2018) identified an optimal solution with respect to total travel cost for one meetup site location along the road network, examining POIs within a search radius. Unfortunately, the model focuses only on a single time, single meeting location, ignoring the possibility of meeting sequentially at different locations. Optimizing sequential meetup locations is not the simple combination of individual optimal meetup locations since choices influence other options. Moreover, individual behavior highlights that travel cost is not the only consideration to selecting a good meetup location. Beyond reducing travel cost, accounting for attendee preference is also important in this research. Integrating cost and preference requires a bi-objective approach to sequential activity location selections. There may be different and/or conflicting goals for the optimization task. For example, the café located at an "optimal" location may have low preference evaluation while a highly regarded café may be too far away.

A multi-objective trajectory optimization (MOTO) model is now introduced. Consider the following notation:

n = number of attendees
i = index of attendees ($i = 1 \dots n$)
o_i = origin of attendee i
e_i = destination of attendee i
T = total number of meetup periods
t = index of meetup period ($t = 1 \dots T$)
j = index of potential meetup locations (also j')
Ω_t = set of potential meetup locations for period t
r_{ij} = preference rating of meetup location j for attendee i's
u_t = the rating requirement for activity in period t
$d_{jj'}$ = shortest distance (or cost) for traveling between meetup locations j and j'

$$Y_{jt} = \begin{cases} 1 & \text{if location } j \text{ is selected for activity in period } t \\ 0 & \text{otherwise} \end{cases}$$

$$Z_{ijj't} = \begin{cases} 1 & \text{if attendee } i \text{ travels from meetup loction } j \text{ in period } t \text{ to } j' \text{ in period } t+1 \\ 0 & \text{otherwise} \end{cases}$$

In each period, the type of activity is known in advance, even though the locations of activities have not been selected. Ω_t is the list of POIs of a certain category regarded as potential meetup locations for period t. r_{ij} is the rating value (binary or integer) based on attributes of POIs, including star level, number of check ins/reviews, price range, ambience, pet friendliness, parking, kid friendliness, etc. One may notice that the shortest network distance between any pair of locations ($d_{jj'}$) is used in this model to describe travel cost, but may reflect any associated measure of spatial interaction. The MOTO is formulated as follows

$$\text{Min} \sum_{i=1}^{n} \sum_{j \epsilon \Omega_1} Y_{j1} * d_{o_i j} + \sum_{i=1}^{n} \sum_{j \epsilon \Omega_t} \sum_{j' \epsilon \Omega_{t+1}} \sum_{t=1}^{T-1} Z_{ijj't} * d_{jj'} + \sum_{i=1}^{n} \sum_{j \epsilon \Omega_T} Y_{jT} * d_{je_i} \tag{1}$$

$$\text{Max} \sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{j\epsilon\Omega_t} r_{ij} * Y_{jt} \tag{2}$$

Subject to:

$$\sum_{j\epsilon\Omega_t} Y_{jt} = 1 \quad \forall \; t \tag{3}$$

$$1 + Z_{ijj't} \geq Y_{jt} + Y_{j't+1} \quad \forall i, \; j \; \epsilon\Omega_t, \; j'\epsilon\Omega_{t+1}, \; t = 1\ldots T-1 \tag{4}$$

$$Z_{ijj't} = \{0{,}1\} \quad \forall \; i, \; t, \; j, \; j' \tag{5}$$

$$Y_{jt} = \{0{,}1\} \quad \forall t, \; j\epsilon\Omega_t \tag{6}$$

Objective (1) seeks a minimal total cost/distance/travel time from origin to the first activity location, then the connection between activity locations and at last to destination for all attendees. Objective (2) is to maximize the total rating value for all of meetup locations. Constraints (3) require exactly one POI to be selected for each activity. Constraints (4) track that each attendee would only travel to the POIs which are selected as meetup locations for activities. Constraints (5) and (6) impose binary restrictions on decision variables. When T is equal to 1, the model without objective (2) can be reduced to the single meetup optimization model detailed in Wang et al. (2018).

Two well-known approaches, the weighting and the constraint methods, are widely used for solving multi-objective and multi-criteria optimization problems. Using the weighting method (e.g. Gass and Saaty, 1955; Zadeh, 1963), a weight is introduced for each objective and subsequently solved as a total weighted sum of the objectives, often where the sum of weights equals to 1. Limitations of the weighting method include the need to search across an infinite range of weighting possibilities as well as the fact that some Pareto optimal solutions cannot be found (see Censor, 1977; Medrano and Church, 2014). An alternative is the constraint method, where only one of the objective functions is considered with the others converted into constraints (Chankong and Haimes, 1983a, 1983b; Haimes et al., 1971). This method works for both convex and nonconvex problems. This paper employs the constraint method to handle multiple objectives.

While it is possible to solve some problem instances exactly, heuristic methods are important for a variety of reasons. As mentioned in the background session, some heuristic algorithms have been used to decrease computation time by narrowing down the number of potential activity locations. In this study, there are cases where the computational cost is extremely high or even prohibitive with the increasing number of attendees and complexity of network. An R* tree-based spatial indexing heuristic is utilized for problem solution. This basic approach has proven effective in the single optimal meetup location case, as reported in Wang et al. (2018), where the candidate datasets are constructed by querying the bounding box of sequential activities using the R* tree spatial index (Beckmann et al., 1990) for POI datasets. Using spatial query with multiple constraints can reduce the search space for finding the sequential multi-meetup location candidates. The minimum bounding geometry reduces the search space for activity location query because only the POIs within the reduced search space are considered. Pseudo code for the heuristic is as follows:

Algorithm: sequential activities optimization (assume $T = 3$)
Input: attendee origins o_i and destinations e_i, requirements for activity in different time periods u_1, u_2, u_3, POI Datasets $\Omega = \Omega_1 \cup \Omega_2 \cup \Omega_3$
Output: optimal activity locations
$\left\{\Omega_1^{R^*},\ \Omega_2^{R^*},\ \Omega_3^{R^*}\right\} = $ R* tree index $(o_i, e_i,\ \Omega)$
MinCost $\leftarrow$initial big value
for $j \in \Omega_1^{R^*}$ **do**
  if $r_{ij} \geq u_1$ for all i **then**
    for $j' \in \Omega_2^{R^*}$ **do**
      if $r_{ij'} \geq u_2$ for all i **then**
        for $j'' \in \Omega_3^{R^*}$ **do**
          if $r_{ij''} \geq u_3$ for all i **then**
            $Cost = Shortest\ Path(o_i, e_i, j, j', j'')$
            if MinCost$>Cost$ then
              MinCost $= Cost$
              $optimal\_loc1^{u1} = j$
              $optimal\_loc2^{u2} = j'$
              $optimal\_loc3^{u3} = j''$
            end if
          end if
        end for
      end if
    end for
  end if
end for
return MinCost, $optimal_{loc}1^{u_1}$, $optimal_{loc}2^{u_2}$, $optimal\_loc3^{u3}$

The algorithm loops through different objective preference weightings in order to consider multiple activity locations. The rating requirement u could be varied for different activities. Attendee preferences are accounted for with respect to minimum travel cost. It is computationally infeasible using a personal computer to derive the all-pair shortest path distance matrix for a large road network for Phoenix. Therefore, the detailed solution approach derives shortest-path travel costs after activity locations are filtered and selected. The ShortestPath function used here is a bi-directional Dijkstra algorithm with binary heaps (see Zhan and Noon, 1998).

Case study

Planning for the identification of a sequential series of activities over space and through time was carried out for a group of people in Phoenix, Arizona, primarily focusing on optimizing travel costs and enhancing social preferences. The study area is approximately 9071 square miles in size. The geographic data layers used for the model included land parcels and roads acquired from OpenStreetMap. The POI layers used in this study were downloaded from the Yelp Dataset Challenge.[1] Yelp is well known as a local-search service powered by crowd-sourced reviews. The Yelp Dataset Challenge provides a subset of its business information

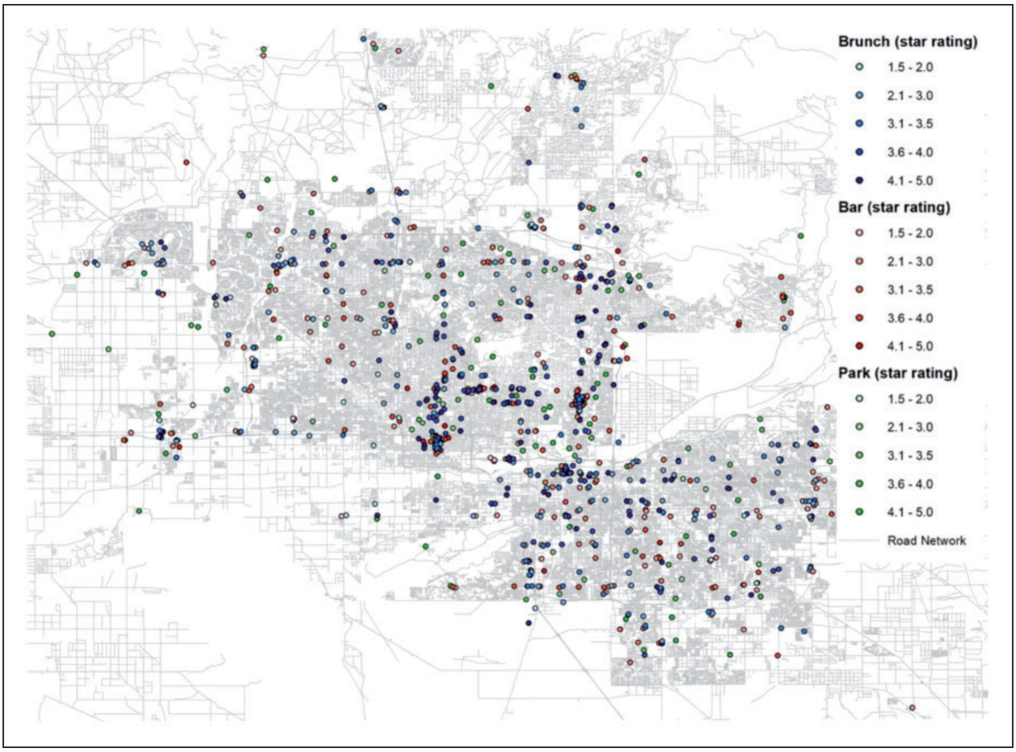

Figure 2. Distribution of potential locations for brunch, hiking, and nighttime entertainment in the study area.

(e.g. hours of operation, service type, attributes, etc.) and review details (e.g. number, rating star, etc.) to researchers. High-resolution imagery for this area was obtained through Google Maps, facilitating identification of POIs. This dataset, along with site visits, enabled verification to enhance positional accuracy, particularly the POI locations and the road network. The study area contains 217,174 road segments and 156,406 nodes after the data cleaning process. Three activities, having brunch, going for a walk, and enjoying evening entertainment in a bar, are considered in the different cases for different groups of attendees in this paper. The origins and destinations of attendees can be the same (e.g. going from and back to their homes), though this was not the case in the scenarios considered here. Figure 2 depicts the distribution of restaurants, parks, and bars, all of which are regarded as potential activity locations. A variety of attributes are used to characterize attendee preferences and requirements. Price range (from 1 to 3), suitability for children, and customer rating scores (from 1 to 5) are used for this study. In Figure 2, the color hue represents different location types and color lightness indicates the rating star level of the corresponding activity sites, with lighter (darker) shades representing a lower (higher) value.

The computational processing was carried out on an Intel(R) Core(TM) i7 2.5 GHz computer running Windows 7 with 8 GB of RAM. ArcGIS was utilized for data creation, management, manipulation, analysis, and display. Xpress, a commercial optimization package, was employed to solve one and two-person scenarios, giving guaranteed optimal results. In scenarios involving more than two people, the R* tree-based spatial indexing heuristic was developed and implemented in C++.

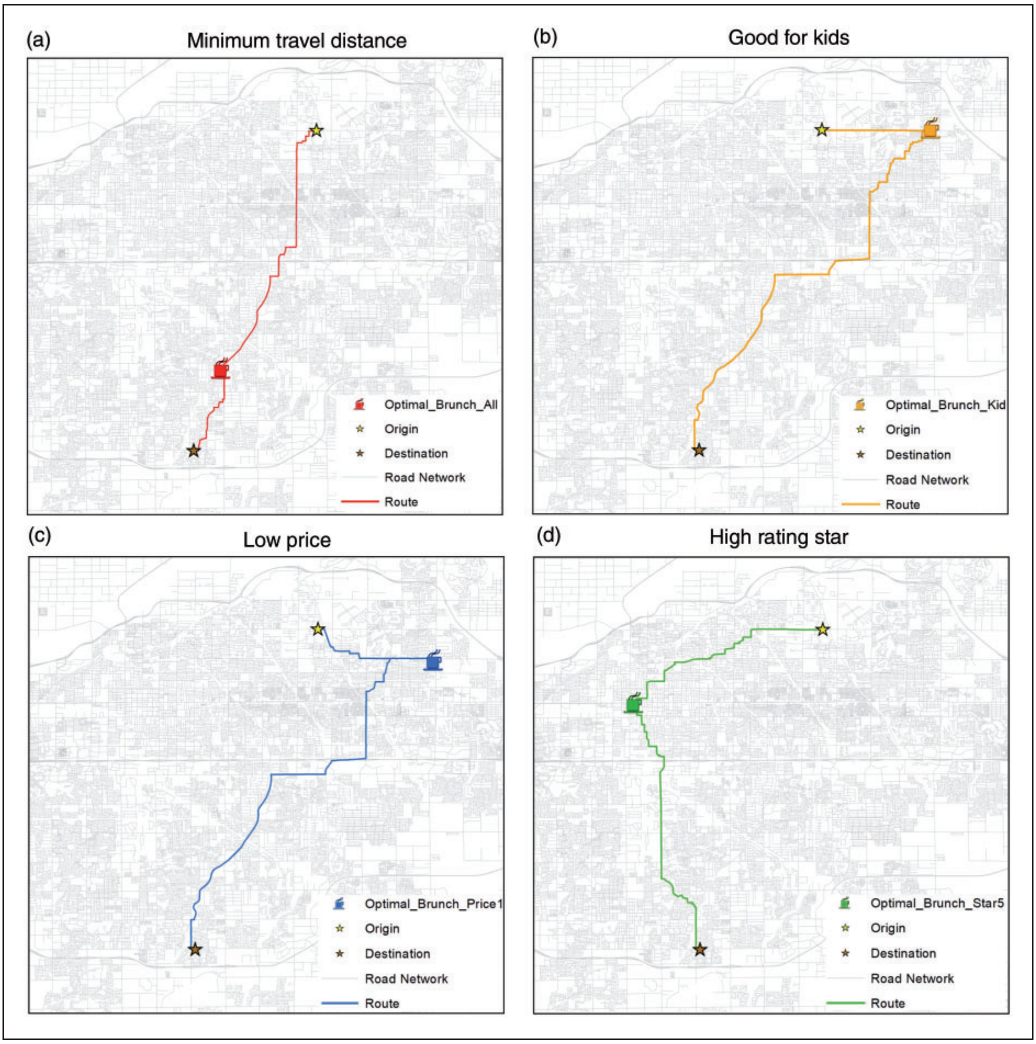

Figure 3. Optimal locations for brunch considering different preferences (a) without additional constraints; (b) good for children; (c) have low price; (d) highly rated.

Scenario A

In this scenario, a single person is considered, and s/he wishes to have brunch. Two cases are examined, with and without personal preference for POI amenities. The person's origin and destination are fixed, represented as yellow stars in Figure 3. Objective (1) is given priority, with the optimal trajectory shown in Figure 3(a). Figure 3(b) to (d) depicts the optimal trajectories when preferences for potential restaurants are taken into account. In this case, consideration for whether they are good for children, have low prices (price range equal to 1), and are highly rated (equal to 5), are explicitly desired. The four trajectories and selected restaurants are significantly different. The optimal restaurant, which minimizes travel distance, is located close to the straight-line connecting the origin and destination, having rating 2, price range 2, and considered not good for children. Compared with the shortest

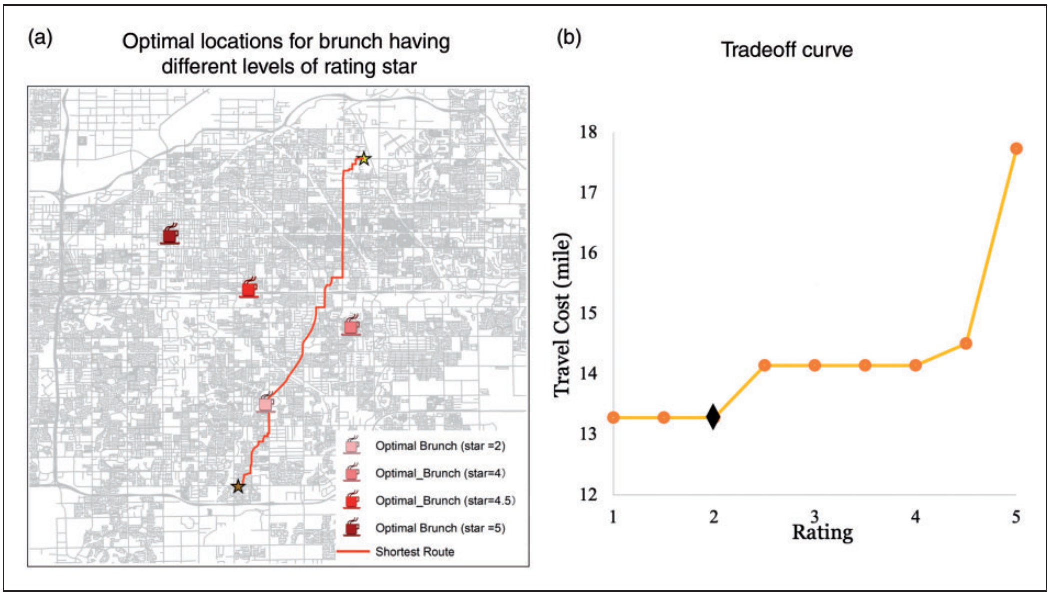

Figure 4. Tradeoff relationship between travel cost and rating constraints. (a) Optimal locations for brunch having different levels of rating star; (b) Tradeoff curve.

distance case (13.27 miles), the traveling cost increases 46.66%, 43.97%, and 29.06% in other three cases. In order to account for restaurant preferences, an attendee has to sacrifice travel cost to some extent. Worth exploring as well is attendee preference, which leads to a tradeoff of sorts between maximizing rating and minimizing travel cost as shown in Figure 4. Along the x-axis is the level of rating star (one option for objective formulation (2)), ranging from 1 to 5 with an interval of 0.5. The y-axis indicates the minimum travel cost, corresponding to objective (1). All of the vertexes along the tradeoff curve are optimal solutions, and depend on the preference given to different objectives. The black diamond in Figure 4(b) is associated with the pink restaurant in Figure 4(a) (also shown in Figure 3 (a)) located on the shortest path. Preference toward a higher star rating means that travel cost will increase, ranging from a low of 13.27 miles to a high of 17.13 miles.

Scenario B

In this scenario, a single person is again considered, but in this case, s/he participates in three activities sequentially. With the location-based service, Yelp, restaurants (brunch), parks and bars are to be accessed in the first, second and third periods, respectively. The distance of the shortest path trajectory (depicted in Figure 5) is 13.93 miles without any constraints on activity preferences. Compared with the situation of only attending brunch in Figure 3(a), the additional activities now favor a brunch location that is closer to the first half of the route. Optimizing sequential activities is not merely a combination of each best meetup location, but rather the activities are balanced simultaneously in order to account for preferences and travel cost.

Benchmark experiments for scenarios A and B are derived to enable comparison of running time and solution accuracy using the developed R*Tree search and a brute force algorithm (i.e. the exhaustive search). Based on 100 sample meetup experiments, the results

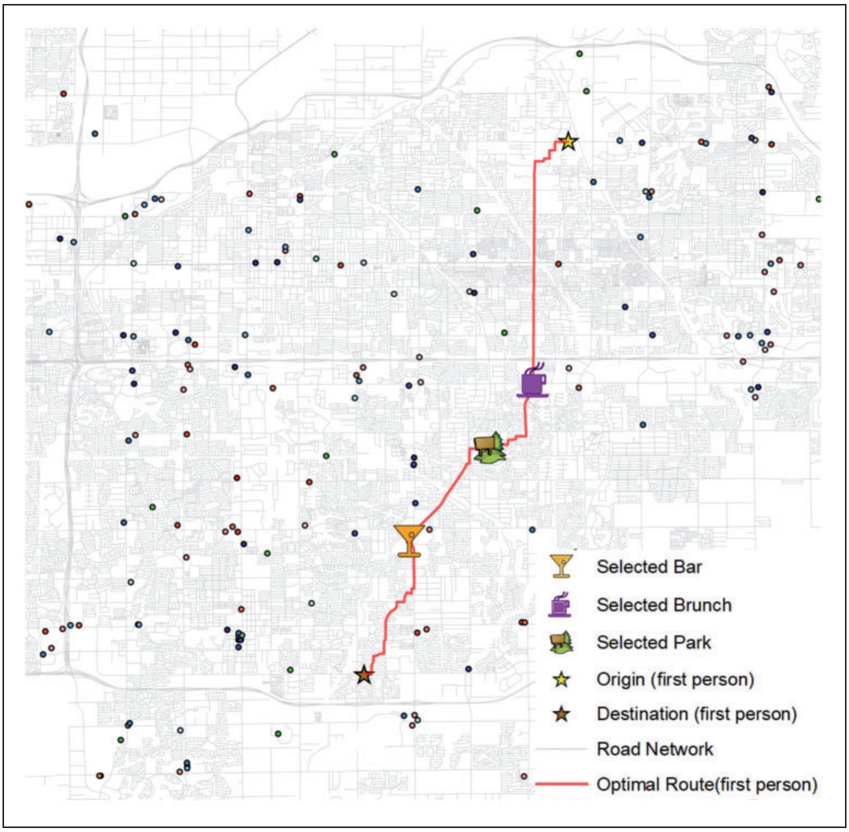

Figure 5. Optimal sequential activity locations and travel routes for one attendee.

Table 1. Benchmark experiment results.

Experiment	Average time—brute force algorithm (s)	Average time—R* Tree Search (s)	Number of optimal solutions
Scenario A	62.912	23.508	96
Scenario B	65.709	24.537	94

in Table 1 show that the average time of R*Tree search is significantly less than that for the brute force algorithm without the R*Tree search, but it cannot guarantee finding the optimal solution. The average time in scenario B is more than in scenario A because the number of constraints in scenario B is larger than in scenario A.

Scenario C

In this scenario, another person joins in the activities, resulting in the three sequential locations shown in Figure 6. All of the activity locations change. The travel cost increases 17.90% compared to that shown in Figure 5. This route and meetup selection of activities is

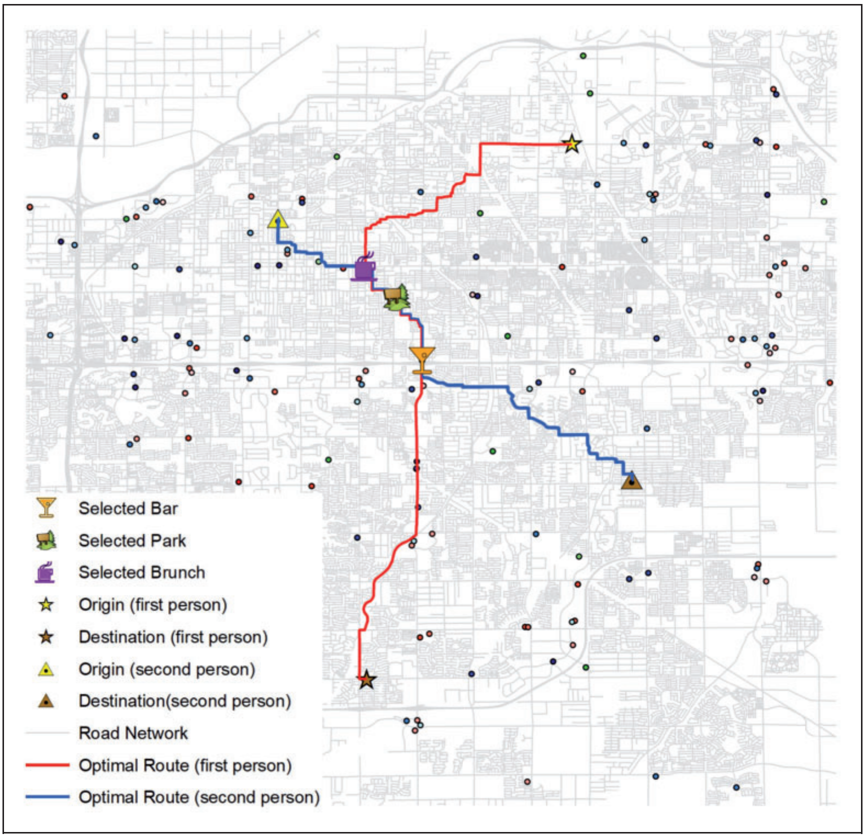

Figure 6. Optimal sequential activity locations and travel routes for two attendees.

a compromise among participants in order to minimize total travel costs and maximize total activity preferences. For the two-person case, the computational time for Xpress to optimally solve the associated problem is around 5 s, including problem input, initialization, and solution.

Scenario D

In this scenario, there are 10 people dispersed throughout the Phoenix metropolitan area who wish to coordinate three sequential activities. For display purposes, the origin and destination are the same, without loss of generality. Without any limitation on the characteristics of the POIs, the three activity locations tend to be congregated close to the center of all the attendee origins/destinations (Figure 7(a)). Figure 7(b) depicts the optimal trajectory and activity locations when attendees call for all activity locations to have a five-star rating. With this extra constraint, the optimal locations are updated and the difference between travel distance for the 10 attendees ranges from –1.2% to 300%, with an average of 82.3%. Compared to the shortest trajectories, an addition of almost 80% of travel cost will be spent on average for each attendee when they request locations having the best star rating.

The solution using the R*tree-based spatial indexing heuristics required approximately 30 s of computational effort for the no preference scenario. This reduces the search space to

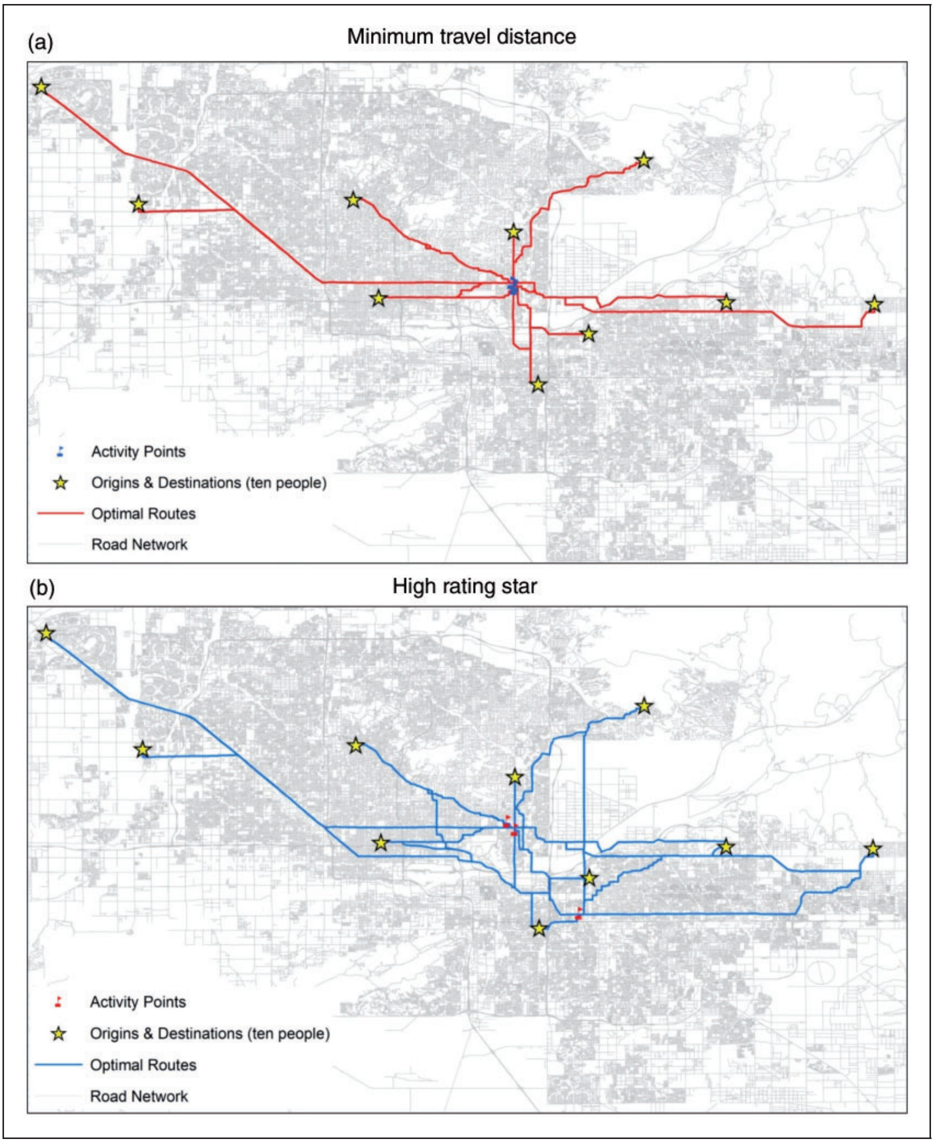

Figure 7. Optimal activity locations and routes for 10 attendees. (a) without any constraints of POIs; (b) with preference of five-start rating of POIs.

335 restaurants, 91 parks, and 366 bars as potential activity locations to be selected for three periods, respectively. When the star rating preference requirement is included, then only 18 restaurants, 8 parks, and 2 bars qualify as activity locations, decreasing the computation time to 7.1 s.

A more general examination of the changes in the objective value with respect to variations in star ratings is explained as follows. The analysis begins with a star level set of constraints using the spatial optimization model. The fluctuation in the travel cost corresponding to different rating requirements may vary based on the distribution of attendees, while sharing a common trend such that if a high rating is required, a high travel cost is possible. Figure 8 summarizes 100 samples where 1 to 5 stars (with a 0.5 star interval) are

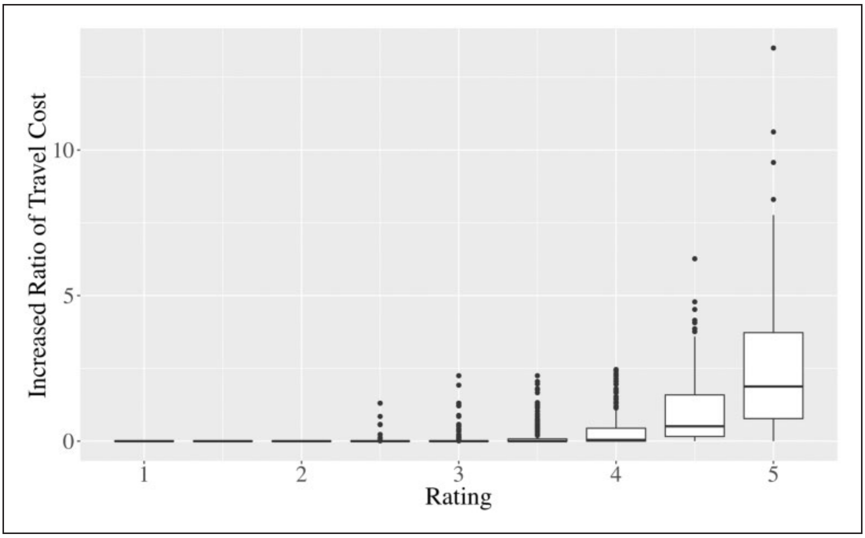

Figure 8. Total travel cost tradeoff with rating constraints (100 samples).

required for each activity location. Each sample has 10 attendees randomly located in the Phoenix Metropolitan area. As the rating requirement increases, the total travel cost gradually increases, especially rises sharply and significantly when rating is higher than 3 stars. Compared with no rating requirement, the average growths of total travel cost are 15.9%, 33.6%, 60.9%, and 135.6%, when rating is expected to be higher than 3.5, 4, 4.5, and 5 stars, respectively. The growth is not apparent when rating is less or equal to 3 stars. One reason could be that large amount of less prominent activity locations are evenly distributed in the study area (as shown in Figure 2). However, there is less chance that a relatively higher star activity location is also the optimal location with minimal travel cost.

Although the global optimal solution is possible in some cases (two or fewer attendees), heuristic solution is desirable for many reasons, including faster solution time as well as eliminating the need for commercial optimization software. Assessment of heuristic solution performance consisted of evaluating 200 cases involving 10 attendees. Of the 200 cases evaluated, 197 were found to be optimal using the R* solution heuristic; the average computation time is reduced by more than 85%, from 159.06 to 21.42 s. In the cases where an optimal solution was not identified using the heuristic, at least one activity location is not within the bounding box of all origins and destinations of attendees. This likely arises when strict requirements (higher star level, more reviews, etc.) for activity locations are imposed since no satisfactory potential POI would be within the search area. A similar test is repeated with the requirement that star levels of activity locations set to be 5. Only 77.5% of cases reached the global optimal solution, reducing the average computation time by approximately 34.33%. The results verify that the decision making for finding a global optimal solution contains a tradeoff between computation time and accuracy. The significance of tradeoff relationship decreases with the additional constraints of people's preference.

Discussion and conclusion

Beyond programming language and computing environment considerations, the number of decision variables and constraints in the optimization model also influence the computation time. For sequential activity selection, if more activities (meetup locations) are needed, the number of decision variables (and computation time) will increase significantly. For example, $Z_{ijj't}$ is the variable determining whether attendee i travels from meetup location j in period t to j' in period $t+1$. The number of $Z_{ijj't}$ variables is $n_i * n_j * n_{j'} * T$, for which n_i, n_j (also $n_{j'}$), and T are the number of attendees, potential activity locations, and total meetup periods, respectively. An interesting issue worth further investigation is how to reduce computation time in activities scheduling with a more expansive regional road network.

External Web services and location-based data streams have the ability to provide geographic context, which influences the travel cost in the spatial optimization model. In a practical engineering implementation, this context information may have a great effect on decision making with respect to activity locations and route planning. The introduction of traffic delay information on road segments will enable context-awareness for finding sequential activity locations under different scenarios. In this problem, the travel time for each road segment, estimated by statistical traffic information, could substitute the shortest distance ($d_{jj'}$) in order to better represent travel cost in the model. In real-world settings, access to available data and contextual information is necessary. As long as these requirements can be structured as constraints, they can be added to the multi-objective optimization model.

In summary, identifying optimal interaction locations with space–time constraints on the road network in order to satisfy human social activity needs is essential, yet there are concerns about travel cost and attendees' preference. This study formalizes the problem of finding sequential activity locations for multiple attendees as a spatial optimization model. A methodology for siting activities is applied relying on the use of context information, GIS, and spatial optimization. Four scenarios were explored for finding optimal activity locations and routes in the Phoenix metropolitan area in order to demonstrate the feasibility and capacity of our framework. The findings show that the optimal activity locations and trajectories change associated with the participation of attendees, the order of activities, the preference of attendees, etc. Attendees' requirements are meaningful and important factors to consider in the decision-making process when preference and context information is important. The proposed method can be applied in many real-world applications such as tourism trip planning, carpooling services, and logistics management.

The generalized model proposed in this paper could be extended in future work. The travel cost objective function in this study accounted for all attendees. However, minimizing the maximum travel cost of any attendee or minimizing only some attendees would be an interesting direction for future research. In addition, different types of POI attributes could be considered for attendee preferences. How to balance the different preferences among all attendees will be another issue that is worth further exploration. Because an attendee may change their mind at or between activities, optimal trajectories could be updated dynamically during the trip. Future research therefore would move toward capabilities for real-time activity coordination and navigation. As computational time increases dramatically with the increase in the size of road network, a parallel computational framework may be useful to enhance efficiency.

Author's note

Shaohua Wang is not affiliated University of California at Santa Barbara.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Dr Shaohua Wang would like to thank the funding support from the National Postdoctoral International Exchange Program of China (No. 20150081). Dr Song Gao would like to thank the support for this research provided by the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison with funding (No. 135-AAC5663) from the Wisconsin Alumni Research Foundation.

ORCID iDs

Xin Feng https://orcid.org/0000-0001-6434-3895
Shaohua Wang https://orcid.org/0000-0001-8651-9505
Song Gao https://orcid.org/0000-0003-4359-6302

Note

1. https://www.yelp.com/dataset_challenge.

References

Aingworth D, Chekuri C, Indyk P, et al. (1999) Fast estimation of diameter and shortest paths (without matrix multiplication). *SIAM Journal on Computing* 28(4): 1167–1181.

Beckmann N, Kriegel HP, Schneider R, et al. (1990) The R*-tree: An efficient and robust access method for points and rectangles. *ACM SIGMOD Record* 19(2): 322–331.

Buchin M, Dodge S and Speckmann B (2012). Context-aware similarity of trajectories. In: *International conference on geographic information science*, Columbus, OH, USA, 18–21 September 2012, pp.43–56. Berlin: Springer.

Censor Y (1977) Pareto optimality in multiobjective problems. *Applied Mathematics & Optimization* 4(1): 41–59.

Chankong V and Haimes YY (1983a) Optimization-based methods for multiobjective decision making—An overview. *Large Scale Systems* 5: 1–33.

Chankong V and Haimes YY (1983b) *Multiobjective Decision Making*. Amsterdam: Elsevier.

Church RL and Murray AT (2009) *Business Site Selection, Location Analysis, and GIS*. Hoboken: John Wiley & Sons.

Clarke K (2011) *Getting Started with Geographic Information Systems*. 5th ed. Upper Saddle River: Prentice Hall.

Cooper L (1963) Location-allocation problems. *Operations Research* 11(3): 331–343.

De Berg M, Cheong O, Van Kreveld M, et al. (2008) *Computational Geometry: Introduction*. Berlin: Springer, pp.1–17.

Dijkstra EW (1959) A note on two problems in connexion with graphs. *Numerische Mathematik* 1(1): 269–271.

Dodge S, Weibel R, Ahearn SC, et al. (2016) Analysis of movement data. *International Journal of Geographical Information Science* 30(5): 825–834.

Gass S and Saaty T (1955) The computational algorithm for the parametric objective function. *Naval Research Logistics Quarterly* 2(1): 39–45.

Hägerstrand T (1970) "What about people in regional science?" *Papers of the Regional Science Association* 24(1): 6–21.

Haimes YY, Lasdon LS and Wismer DA (1971) On a bicriterion formulation of the problems of integrated system identification and system optimization. *IEEE Transactions on Systems, Man, Cybernetics* 1: 296–297.

Kim HM and Kwan MP (2003) Space-time accessibility measures: A geocomputational algorithm with a focus on the feasible opportunity set and possible activity duration. *Journal of Geographical Systems* 5(1): 71–91.

Kuijpers B, Miller HJ, Neutens T, et al. (2010) Anchor uncertainty and space-time prisms on road networks. *International Journal of Geographical Information Science* 24(8): 1223–1248.

Kuijpers B and Othman W (2009) Modeling uncertainty of moving objects on road networks via space–time prisms. *International Journal of Geographical Information Science* 23(9): 1095–1117.

Lombard K and Church RL (1993) The gateway shortest path problem: Generating alternative routes for a corridor location problem. *Geographical Systems* 1(1): 25–45.

Medrano FA and Church RL (2014) Corridor location for infrastructure development: A fast bi-objective shortest path method for approximating the pareto frontier. *International Regional Science Review* 37(2): 129–148.

Miller HJ (1991) Modelling accessibility using space-time prism concepts within geographical information systems. *International Journal of Geographical Information Systems* 5(3): 287–301.

Miller HJ (2005) Necessary space-time conditions for human interaction. *Environment and Planning B: Planning and Design* 32(3): 381–401.

Scaparra MP, Church RL and Medrano FA (2014) Corridor location: The multi-gateway shortest path model. *Journal of Geographical Systems* 16(3): 287–309.

Shaw SL and Yu H (2009) A GIS-based time-geographic approach of studying individual activities and interactions in a hybrid physical–virtual space. *Journal of Transport Geography* 17(2): 141–149.

Siła-Nowicka K, Vandrol J, Oshan T, et al. (2016) Analysis of human mobility patterns from GPS trajectories and contextual information. *International Journal of Geographical Information Science* 30(5): 881–906.

Song Y, Miller HJ, Stempihar J, et al. (2017) Green accessibility: Estimating the environmental costs of network-time prisms for sustainable transportation planning. *Journal of Transport Geography* 64: 109–119.

Wagner D and Willhalm T (2007) Speed-up techniques for shortest-path computations. In: Thomas W and Weil P (eds) *STACS* 2007. Berlin: Springer, pp.23–36.

Wang H, Terrovitis M and Mamoulis N (2013) Location recommendation in location-based social networks using user check-in data. In: *Proceedings of the 21st ACM SIGSPATIAL international conference on advances in geographic information systems*, Orlando, FL, USA, 5–8 November 2013, pp.374–383. New York: ACM.

Wang S, Gao S, Feng X, et al. (2018) A context-based geoprocessing framework for optimizing meetup location of multiple moving objects along road networks. *International Journal of Geographical Information Science* 32(7): 1368–1390.

Ye M, Yin P, Lee WC, et al. (2011) Exploiting geographical influence for collaborative point-of-interest recommendation. In: *Proceedings of the 34th international ACM SIGIR conference on research and development in information retrieval*, Beijing, China, 25–29 July 2011, pp.325–334. New York: ACM.

Zadeh L (1963) Optimality and non-scalar-valued performance criteria. *IEEE Transactions on Automatic Control* 8(1): 59–60.

Zhan FB and Noon CE (1998) Shortest path algorithms: An evaluation using real road networks. *Transportation Science* 32(1): 65–73.

Biographical notes

Xin Feng is currently a PhD candidate at the Department of Geography, University of California, Santa Barbara. Her research interests include geographic information science; spatial optimization; location modeling; spatial representation; urban, regional, and natural resource planning and development; transportation; and techniques to support interactive planning and decision making. She develops and applies spatial optimization approaches to locate and allocate public or private resources in various social planning contexts.

Shaohua Wang received PhD degree in the field of cartography and GIS from the University of Chinese Academy of Science in 2013. From 2013 to 2018, he was postdoctoral fellow at Institute of Geographic Sciences and Natural Resources Research, CAS. From 2016 to 2018, he was a visiting scholar in Geography in the Department of Geography, University of California, Santa Barbara. His major research interests include spatial optimization, spatial cloud computing, and spatiotemporal big data analytics.

Alan Murray is a professor in the Department of Geography at University of California at Santa Barbara. He is editor of *International Regional Science Review*, associate editor and/or editorial board member for a range of international journals, including *Geographical Analysis*, *Transactions in GIS*, *Papers in Regional Science*, *Computers, Environment and Urban Systems*, *Socio-Economic Planning Sciences*, and *Annals of the Association of American Geographers*. His research and teaching interests include geographic information science; spatial optimization; health informatics; urban growth and development; land use planning; urban, regional, and natural resource planning and development; and infrastructure and transportation systems. He is the author of two books and over 230 research articles, book chapters and proceedings papers. He has received a number of awards from the American Association of Geographers, and is the recipient of the Walter Isard Award for Scholarly Achievement from the North American Regional Science Council, among others.

Yuanpei Cao (PhD, University of Pennsylvania) is currently a data scientist at Airbnb Incorporation, San Francisco. His research interests include high-dimensional statistical inference, big data analytics, and recommender system. He utilizes multi-data sources, numerical optimization, and statistical models to develop recommender system in advertising, marketing, and payment industry.

Song Gao (PhD, University of California, Santa Barbara) is currently an assistant professor in the Department of Geography, University of Wisconsin, Madison. His main research interests include place-based GIS, spatiotemporal big data analytics and geospatial semantics. He utilizes multi-data sources, spatiotemporal analyses, and machine learning methods to study human mobility patterns and urban dynamics.