SmartAdP: Visual Analytics of Large-scale Taxi Trajectories for Selecting Billboard Locations

Dongyu Liu, Di Weng, Yuhong Li, Jie Bao, Yu Zheng, Huamin Qu, and Yingcai Wu

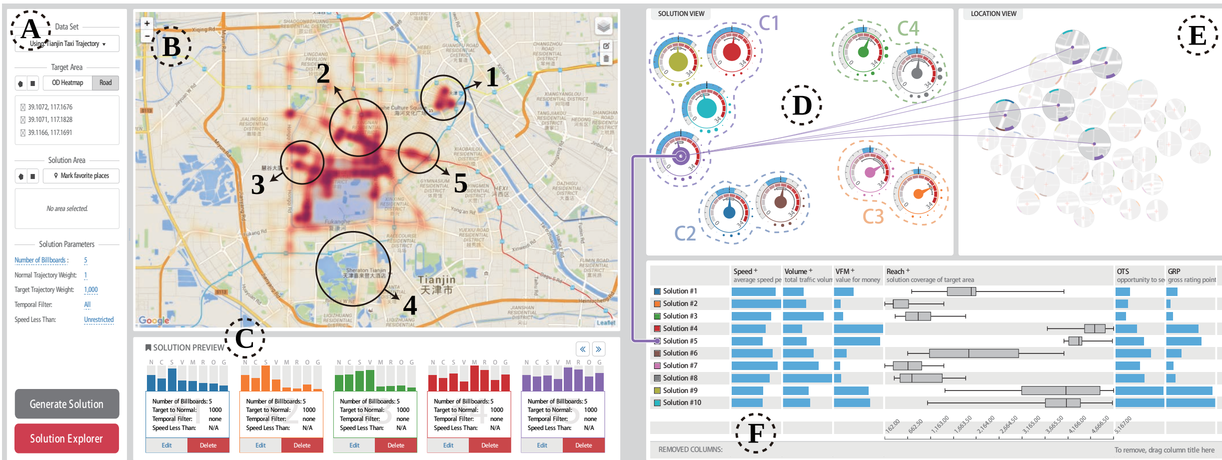

Fig. 1. SmartAdP system. (A) Dashboard View shows the information of the current solution for billboard placements. (B) Map View provides a visual summary of the geospatial environment. (C) Solution Preview lists the parameters and statistics of the candidate solutions. (D) Solution View lays out all the solutions as glyphs to reveal the relationships among the solutions. (E) Location View supports in-depth analysis at the fine-grained location level. (F) Ranking View displays multi-typed ranks of the solutions.

Abstract— The problem of formulating solutions immediately and comparing them rapidly for billboard placements has plagued advertising planners for a long time, owing to the lack of efficient tools for in-depth analyses to make informed decisions. In this study, we attempt to employ visual analytics that combines the state-of-the-art mining and visualization techniques to tackle this problem using large-scale GPS trajectory data. In particular, we present SmartAdP, an interactive visual analytics system that deals with the two major challenges including finding good solutions in a huge solution space and comparing the solutions in a visual and intuitive manner. An interactive framework that integrates a novel visualization-driven data mining model enables advertising planners to effectively and efficiently formulate good candidate solutions. In addition, we propose a set of coupled visualizations: a solution view with metaphor-based glyphs to visualize the correlation between different solutions; a location view to display billboard locations in a compact manner; and a ranking view to present multi-typed rankings of the solutions. This system has been demonstrated using case studies with a real-world dataset and domain-expert interviews. Our approach can be adapted for other location selection problems such as selecting locations of retail stores or restaurants using trajectory data.

Index Terms—optimal billboard locations, taxi trajectory, visual analytics, comparative analysis

1 Introduction

Billboards are the most common forms of outdoor advertising. Despite the decline of other traditional advertising media, billboard advertising remains critically important, because people today are spending considerable time in transit. Billboard advertising has several evident advantages like prominent visibility, low cost per mille, and superior accumulation of local influence over other advertising methods.

However, launching a successful billboard campaign is difficult, because many factors should be cautiously considered, such as content design, locations, and visibility. Among them, the geographical locations of billboards are considered the most critical. Appropriate billboard locations can increase audience exposure significantly, whereas inappropriate ones lead to waste of time and investment. Conventional approaches involve manually computing traffic volume, conducting travel surveys, and inviting experts to build mathematical models [23]. As a result, this process is time consuming, less flexible, and can be completed by professional advertising agencies only. Currently, several famous outdoor advertising companies, such as *APN* [1] and *LAMAR* [2], have provided online campaign planning tools that include billboard planners with comprehensive information of each billboard, such as the traffic statistics and demographics.

These tools enable planners to create and offer various tailored solutions to their business customers; however, several limitations have been identified. First, the acquisition of such comprehensive data is expensive, thereby limiting the data coverage. Second, the large solution space with many advertising factors still poses difficulties for experts

- *D. Liu is with Zhejiang University and the Hong Kong University of Science and Technology. This work was done when Dongyu Liu was a visiting student supervised by Yingcai Wu at Zhejiang University. Email: dliuae@cse.ust.hk*
- *D. Weng, and Y. Wu are with the State Key Lab of CAD & CG, Zhejiang University. Yingcai Wu is the corresponding author. E-mail: ycwu@cad.zju.edu.cn, mystery.wd@gmail.com*
- *Y. Li is with University of Macau. E-mail: yb27407@umac.mo*
- *J. Bao and Y. Zheng are with Microsoft Research, Beijing, China. E-mail: jiebao@microsoft.com, yuzheng@microsoft.com*
- *H. Qu is with the Hong Kong University of Science and Technology. E-mail: huamin@cse.ust.hk*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx/

in utilizing the tools to create proper solutions without leveraging the computational power of machines. Third, in most cases, several candidate solutions need to be provided to customers. Since customers' own criteria and expectations may vary significantly from one person to another, they are in urgent need of tools that can help quickly distinguish the commonalities and differences among multiple solutions. Unfortunately, existing tools cannot support this kind of analysis.

In this study, we attempt to employ visual analytics techniques to overcome these limitations using taxi GPS trajectory data. We utilize the taxi trajectory data for two reasons: (1) the data is relatively easy to collect and provide citywide coverage in most cities; and (2) the large-scale data can effectively reveal the underlying traffic patterns [8, 49]. To our knowledge, previous studies have not reported the use of taxi trajectory data to address the billboard location selection and solution-level comparison problems.

However, there are two major challenges in visual analytics of the trajectory data: (1) dealing with a wide solution space to determine the desired solutions; (2) creating an intuitive visualization to facilitate comparative analysis. A typical billboard solution involves a set of locations. The solution space is almost infinite because of numerous possible billboard locations in a city. Thus, the cost of searching for an optimal solution under multiple criteria is prohibitively expensive. Furthermore, a billboard location can be characterized by multiple spatio-temporal attributes, such as traffic volume, speed, origins and destinations (OD), and surrounding environment (e.g., points of interest (POIs)). Visual comparisons among solutions can be viewed as comparing different groups of locations, which are depicted by multiple spatio-temporal attributes. Hence, creating a concise and readable visual representation to facilitate comparison is non-trivial.

To deal with the huge space, we tightly integrate the knowledge and expertise of humans with the computational power of machines. In particular, we introduce a novel visualization-driven data mining model based on a tailored application-specific data index mechanism to efficiently generate billboard candidate locations. To enable effective visual comparison, we propose a set of coupled visualizations, allowing users to compare solutions from multiple perspectives and different levels of details. With these techniques, we develop a visual analytics system called Smart Advertising Placement (SmartAdP) to visualize and explore the heterogeneous urban data, such as taxi trajectory data, POI data, and a citywide geospatial route network. Our system can also be easily adapted to other applications with respect to recourse allocation, such as retail chain location selections.

The major contributions of our study are as follows:

- A systematic characterization of the problem of billboard location selection using taxi trajectory data, and a thorough discussion and summary of the design requirements and space.
- An interactive framework to generate billboard solutions with a novel visualization-driven data mining model and a tailored application-specific data index mechanism.
- A set of new visualization techniques to empower end users to explore the major features of multiple solutions and compare the commonalities and differences among them.

2 Related Work

This section discusses the prior studies closely related to our work.

Trajectory query processing has received considerable attention in recent years. The queries on trajectories can be categorized into three types, namely, the point, region, and trajectory query [12, 48]. Point query aims to find the points with an expected spatio-temporal relationship to several given trajectories or retrieve the trajectories that have a specific relationship with a few given points. Examples include finding the k nearest neighbors to a path in road networks [9], or querying for the nearest trajectories to a point [21]. Similar to point query, region query seeks the trajectories within the specified spatio-temporal regions or identifies the regions with a specific relationship to a set of given trajectories. Examples include retrieving the most frequent path between a user-specified OD [42], as well as discovering the gathering patterns from trajectories [46]. Trajectory query focuses on searching the trajectories that possess similar features within a given trajectory set (i.e., classification or clustering [28, 29]). Our query can be categorized as point query, which aims to find points (i.e., billboard locations) within user specified regions that can cover maximum trajectories. In our application scenario, users aim to highlight important regions or trajectories while playing down other regions or trajectories. Hence, the query method should allow users to flexibly set different weights for different regions or trajectories. However, identifying optimal locations with millions of trajectories in real time is challenging due to the large search space and high algorithm complexity. To our knowledge, the problem has not been systematically studied in the areas of database and data mining. In this study, we present a novel visualization-driven data mining model to support this kind of query, in which users are allowed to participate in the improvement of the quality and efficiency of query processing.

Billboard location selection is one of multicriteria decision making (MCDM) problems in a spatial context [33, 34]. Before making a choice, billboard campaign planners typically have to run appropriate trade-offs among multiple conflicting criteria. Karamshuk et al. [22] demonstrate the power of computational methods (i.e., data mining) to tackle the problem of optimal retail store placement by using human mobility information (i.e., Foursquare check-in dataset) and POIs. However, sometimes the optimal solution cannot be solely generated by computer, as people always have their own opinions and criteria. Exploratory visualization methods are commonly used to address the problem and make a well-substantiated decision. As such, several mechanisms of integrations of computational methods with interactive visualization tools by multiple coordinated views have been suggested [5]. Nonetheless, to our knowledge, there is little work on combining data mining and visualization techniques to address the problem of billboard location selection with taxi trajectory data.

Taxi trajectory visualization has been extensively studied and applied in traffic monitoring [43], mobility pattern discovery [16], route recommendation [30], urban planning [19], and so on. Different from the prior studies, our research focuses on solving the billboard location selection problem by utilizing visual analytics techniques. In this light, many trajectory visualization techniques [7, 8, 31] can be used.

Taxi trajectories can be viewed as a sequence of time-ordered spatial points with multiple attributes. By leveraging point-relevant visual channels and animation techniques, trajectories can be intuitively observed [38]. With the increase of points, one can use a heatmap to show the integrated quantity of points in a map. Besides, researchers propose several line-based aggregation visualization techniques to depict traffic flow in a distributed network, such as density map [39] and aggregation flow [6]. These techniques are capable of revealing movement patterns intuitively. To depict both spatio-temporal information and related attributes of taxis (speed, direction, and volume, etc.), several space-time-cube based and radial metaphor based visualizations are also introduced [30, 41]. However, these techniques are not designed for visual comparison of complex advertising solutions with dozens of billboards that possess special features, such as geolocations, cost, and coverage. Therefore, these techniques cannot be simply applied in our solution comparison task.

Visual comparison is one of the most fundamental and common visualization tasks [24]. There are three widely-accepted categories of visual comparison: juxtaposition (i.e., side by side), superposition (i.e., overlay), and explicit encoding (i.e., visually showing differences or correlations) [17]. Various visualization approaches have been developed on the basis of these three basic methods for different comparison tasks [40, 44]. For example, VisLink [10] links the same objects in two juxtaposed visualizations. Kehrer et al. [25] use hierarchically organized small-multiple displays (i.e., juxtaposition) to compare multi-variate data comprising categorical and numerical information. They also support superposition and explicit linking in a small-multiple display cell. Our work adopts a juxtaposition method, allowing users to conduct multi-level and multivariate comparative analysis. The work unifies three views, namely, the solution, location, and ranking view, through explicit visual linking and user interactions (Fig. 1(D, E, F)).

3 BACKGROUND

This section introduces the background on billboard advertising and the types of data used. Thereafter, the analytical tasks are discussed.

3.1 Background Knowledge

Billboard location selection is a multidisciplinary research problem that involves advertising, communication, and urban computing. In the past year, we have been working with three experts in these fields. In particular, one expert is a manager from an advertising agency who has considerable experience in advertisement planning (EA), another one is a postdoctoral researcher in communication (EB), and the third is a senior researcher in urban computing (EC). EB and EC were invited specially to solve the billboard location selection problem that was originally proposed by EA.

As with the real estate business, billboard locations are considered a decisive factor for a billboard campaign. However, different people may have different opinions on locations. We run structured interviews with EA for several rounds and summarize the main challenges for billboard location selection as follows.

◇ **Finding befitting areas to place billboards.** The first step is to determine several areas to place billboards based on customers' requirements. Areas frequently visited by the target audience are desired. However, this type of information is difficult to gather. Thus, planners from outdoor advertising companies often make recommendations based on their own experience and knowledge.

◇ **Selecting proper locations in the specified areas.** Each specified area contains numerous locations for placing billboards. Planners have to spend considerable time on manually selecting proper locations from the locations in each area based on the basic information of each location; such information includes, the surrounding POIs, daily circulation, and cost. The best and worst locations are easy to determine, but the ones in between are difficult to distinguish. Moreover, the information on candidate locations is often incomplete. Furthermore, information on passers-by along the locations is rather rare. Planners often obtain the information through field studies or based on their own experience and knowledge.

◇ **Evaluating a solution and convincing customers.** Assessing a solution still remains difficult. Field studies, such as those involving the use of questionnaires, are usually conducted. Nevertheless, the limited data and lack of appropriate tools have resulted in difficulty in convincing customers who often have different criteria.

◇ **Providing customers with multiple solutions.** Planners have to formulate multiple solutions and present them to customers. This process frequently costs substantial time. To our knowledge, no tool is available for both planners and customers to further analyze and compare multiple candidate solutions.

The feedback from EA suggests that a visual analytics system is necessary to empower planners to formulate multiple solutions quickly, as well as to compare the solutions effectively. To design this system, we followed a user-centered design process [36] and involved experts in every stage of the iterative development since early 2015.

To elaborate the problem, we formally defined two terms, namely, *target area* and *target trajectory*. *Target areas* often refer to the areas where the target audience lives or works. *Target trajectories* are those whose origins or destinations are within the target areas.

- *Traffic volume* refers to the volume of passengers who are likely to see the billboard.
- *Traffic Speed* refers to the amount of time that passengers allocate to see the billboard.
- *Traffic OD* can reveal a passenger's demographic information to assist in identifying the target trajectories.
- *Environment* reveals the information on the surrounding POIs and the residents living or working around the location.
- *Cost* indicates the cost of a billboard.

We adopted three widely-accepted performance indicators [23] for our scenario based on the suggestions of the domain experts.

- *Coverage/Reach* represents the percentage of the covered target trajectories among all target trajectories. A trajectory is covered if it passes by at least one of the billboards (i.e., at least one contact) within a specified time.
- *Opportunities to see (OTS)* indicates the average number of billboard contacts among all **target trajectories** that see a billboard of the campaign.
- *Gross rating points (GRP)* measures the average number of billboard contacts that 100 **target trajectories** produce ($GRP = reach * OTS * 100$).
- *Value for money (VFM)* states the value of the cost ($VFM = covered\ target\ trajectories/total\ cost$).

3.2 Data Abstraction

We mainly used three types of data collected in one city, where taxis are a common form of transportation. The detailed information is described as follows:

Road network data comprises $133,726$ road segments (the average length is 243 m) and $99,007$ vertices (i.e., intersections of road segments) in the city.

GPS trajectory data includes the trajectories of $3,501$ taxis from the city in two months. A total of $3,500$ sample GPS points are collected for each taxi in one day at a sampling rate of 24 seconds per point. These points constitute approximately 4 million trajectories (segmented by passenger on/off events).

POI data contains $154,633$ points in the city. Each POI is denoted by its ID, category, and GPS location.

3.3 Task Analysis

By discussing with the experts in the form of structured interviews, we compiled a list of analytical tasks.

R.1 **Spatio-temporal distribution:** *How are the target trajectories distributed across the city? What is the difference between weekday and weekend?* This information help users select the target areas judiciously.

R.2 **Location recommendation:** *How many billboards should be placed in the target areas? Where are the optimal locations?* Manually selecting proper locations without computational aids is time-consuming and may easily lead to suboptimal solutions. Therefore, an interactive visual exploratory tool combining with auto recommendation mechanisms (i.e., computational methods) is strongly required.

R.3 **Location assessment:** *How good is a billboard location? Why is it selected for a billboard?* Users want to easily access the detailed information of a location, such as its cost, neighboring environment, traffic volume, speed, and OD, to enable them to make informed decisions.

R.4 **Solution assessment:** *How effective is a billboard solution? How does it satisfy customers' requirements?* Users want to know the performance indicators of the selected solution such as reach, traffic volume, traffic speed, and cost, to furhter estimate whether it can meet the customers' requirements. In particular, the geospatial distribution of billboards should also be provided.

R.5 **Solution comparison:** *What are the differences and similarities among multiple candidate solutions?* An in-depth understanding of the differences and similarities among the candidate solutions can help planners elaborate the selected solutions and explain their choices to customers.

R.6 **Solution classification:** *How many groups of candidate solutions exist? How these groups differentiate from each other?* Users are able to formulate multiple solutions in short time with the assistance of computational models. Thus, it is critical to know how the generated solutions can be grouped to obtain a quick overview of the solutions.

R.7 **Solution ranking:** *What is the ranking of multiple solutions? Which ones are optimal?* Users may have different opinions on the optimal, thereby an interactive ranking method should be provided, allowing users to rank solutions as desired. A reasonable ranking can help planners quickly find the desired solutions.

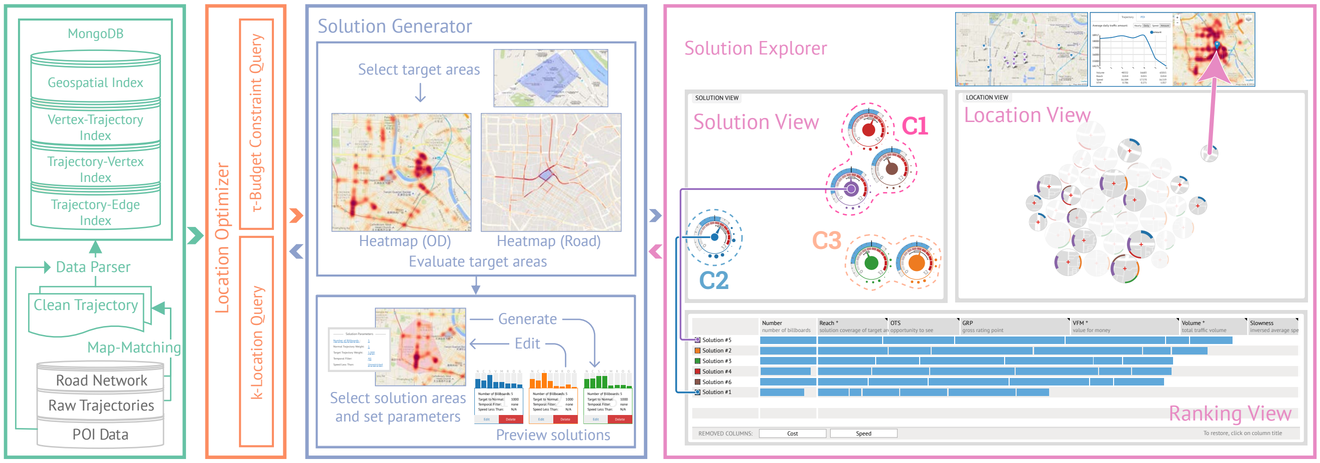

Fig. 2. SmartAdP comprises three major components: data manager, location optimizer, and visualization explorer (i.e., solution generator and explorer). The raw data is preprocessed and stored into the four application-specific data indexes for the use of location optimizer. The location optimizer aims to determine the optimal billboard locations by leveraging the computational power of machines. The solution generator interacts with the optimizer to create good candidate solutions. Meanwhile the solution explorer enables visual exploration and comparison of these solutions.

4 SYSTEM ARCHITECTURE

SmartAdP is a web-based application developed under the full-stack framework of MEAN.js (i.e., MongoDB, ExpressJS, AngularJS, and Node.js). The visual analysis module is implemented using D3.js and Leaflet.js. We deployed the back-end part into our server with 2.40GHz Intel Xeon E5-2620 CPU and 64GB memory. Fig. 2 shows the SmartAdP's system architecture.

The solution generator helps users formulate a candidate solution. Users need to select several target areas initially, and then two types of heatmaps are provided to help users determine the befitting solution areas to place billboards (R1). When the solution areas are determined, users set the parameters of model and obtain a recommended solution from the location optimizer (R2). Meanwhile, users can assess whether the selected locations or the generated solutions are good enough (R3, R4) and make adjustments accordingly. To further explore and compare multiple solutions, users can switch to solution explorer that comprises three sub-views. The solution view shows a high-level overview of the basic information of each solution and the relationships among the solutions (R5, R6). The location view further assists users in identifying the relationships at a locational level (R5, R6). The ranking view visualizes the detailed performance related to the attributes of each solution (R4, R5, R7).

5 MODEL

This section first describes the construction of data structure and then introduces our novel visualization-driven model.

5.1 Data Structure Construction

The major data structures utilized in this study are trajectory-edge, trajectory-vertex, and vertex-trajectory indexes. We refer to the road segment in the road network as the edge, and the intersection (i.e., candidate location for placing billboard) as the vertex.

Trajectory-Edge Index, I_{te}. A GPS trajectory is a sequence of time-ordered spatial points. We firstly apply a map-matching algorithm [32] to map the spatial points of a given trajectory to the underlying road network. Thereafter, the trajectory-edge index can be constructed. From this index, the road segments passed by a given trajectory can be identified.

Trajectory-Vertex Index, I_{tv}. The trajectory-vertex index records the covered vertices of all trajectories and can be easily constructed from I_{te}. Thus, each entry in I_{te} represents a unique trajectory, which is identified by the trajectory ID Tr_i. $I_{tv}[Tr_i]$ lists all vertices that are passed by Tr_i, that is, $\{\underline{Tr_i} \mid v_i, v_j, \dots\}$.

Vertex-Trajectory Index, I_{vt}. This index is an inverted trajectory-vertex index, where each entry is identified by a vertex v_i on the road network. The entry $I_{vt}[v_i]$ in this index stores all the trajectories that are covered by v_i, that is, $\{\underline{v_i} \mid Tr_i, Tr_j \dots\}$. One can easily obtain the coverage of a given vertex by using this index.

With these indexes, we can directly calculate the statistics (e.g., volume and speed) for each road segment and each location through a series of summation and averaging operations. The reach, OTS, GRP, and VFM can be calculated in the same manner, where only the target trajectories are counted. Besides, a geospatial index can be naturally supported by MongoDB [3]; for example, the vertices or the trajectories (OD) within multiple polygonal regions can be queried.

5.2 Extracting the Optimal Locations

SmartAdP provides two types of interactive queries, namely, k-location query and τ-budget constraint query, to assist the domain experts in selecting the billboard locations. These two queries are based on different scenarios. In particular, the k-location query aims to extract k locations from the candidates, where the cost of billboards is not considered. Meanwhile, the τ-budget constraint query focuses on mining a set of locations with the total cost not exceeding the budget τ. Each trajectory has different weights in both queries. We define the *coverage value* as the sum of the weights of all the covered trajectories. Thus, both queries are aimed at extracting a set of locations with the maximum coverage value to achieve a satisfactory advertising effect. However, identifying k locations with the maximum trajectory coverage is an NP-hard problem, which is computing infeasible for large k. In this situation, a tradeoff between the efficiency and effectiveness should be considered. Therefore, we finally propose a visualization-driven mining model that not only human knowledge can play a role, but also an efficient searching and pruning strategy is employed (i.e., the greedy heuristic method).

5.2.1 k-Location Query

The literature [15] has proven that the greedy heuristic is the most effective polynomial solution to our problem and can provide $(1 - 1/e)$ approximation to the optimal solution. Algorithm 1 shows the pseudo code of the greedy heuristic for the k-locations query. This algorithm first assigns the weights to each trajectory based on whether they are target trajectories (cf. line 1-6). For example, the weights of trajectories with their OD within the given spatial regions R_{od} are w_{od} (cf. line 4-6), whereas the weights of the remaining trajectories are set to w_{nor}. Then, the coverage value of each candidate vertex can be calculated by adding the weights of its covered trajectories (cf. line 8). Finally, the greedy heuristic is applied in selecting the k locations (cf. line 10-16). In each iteration, the algorithm contains two steps:
Selection: In this step, the algorithm selects the vertex with the maximum coverage value and put it into the result set (cf. line 11- 12). *Updating:* In this step, the algorithm updates the coverage value of all

Algorithm 1 *k*-location query

Algorithm KLocation(Candidate vertices V_{can}, Trajectory-vertex index I_{tv}, Vertex-trajectory index I_{vt}, OD regions R_{od}, Normal weight w_{nor}, OD weight w_{od}, k)

1: identify all the trajectories covered by $V_{can} \rightarrow TR_{can}$
2: **for** each trajectory Tr in TR_{can} **do**
3:     set $w(Tr)$ to w_{nor}
4: identify all the trajectories that one of its OD vertices located in $R_{od} \rightarrow TR_{od}$
5: **for** each trajectory Tr in TR_{od} **do**
6:     set $w(Tr)$ to w_{od}
7: **for** each vertex v in V_{can} **do**
8:     calculate the coverage value $c(v)$ as $\sum_{Tr \in I_{vt}[v]} w(Tr)$
9: $V_{result} := \emptyset$; $TR_{covered} := \emptyset$
10: **for** $i := 0$ to k-1 **do**
11:     pickup v_{max} in V_{can} with the maximum coverage value
12:     $V_{result} := V_{result} \bigcup v_{max}$
13:     **for** Tr in $I_{vt}[v_{max}] - TR_{covered}$ **do**
14:         **for** v in $I_{tv}[Tr]$ **do**
15:             $c(v) := c(v) - w(Tr)$
16:     $TR_{covered} := TR_{covered} \bigcup I_{vt}[v_{max}]$
17: **return** V_{result}

vertices. Specifically, for each newly covered trajectory Tr in the current iteration (cf. line 13), it can identify the passing vertices by using the trajectory-vertex index, i.e., $I_{tv}[Tr]$. The coverage value of every passing vertex v is updated to $c(v) - w(Tr)$ (cf. line 15).

5.2.2 τ-Budget Constraint Query

We denote the cost of placing a billboard in vertex v as $f(v)$, which is estimated by considering two important factors, namely, the distance from central areas and the traffic volume near the location; with the cost of each central area being predefined, the cost of each billboard location (the building cost is not considered here) can be calculated by interpolation. The τ-budget constraint query attempts to extract a set of locations with the total cost within τ. Therefore, the problem can be modeled as the budget constraint maximum coverage problem, and the better solution between the *small cardinality optimal solution* and *modified greedy heuristic* can also provide a performance guarantee $(1 - 1/e)$ [26]. The algorithm follows two steps:

Small cardinality optimal solution: It identifies the optimal solution with small cardinality. The cardinality of the optimal solution can be set to 1, 2, or 3 according to the requirement of the result quality and response time. Typically, a large cardinality tends to have more response time but better performance guarantee.

Modified greedy heuristic: Different from the greedy heuristic in Algorithm 1, the modified greedy heuristic adds the vertex with the maximum utilization ratio to the result set in each round. This process continues until the total cost reaches the budget constraint τ. The utilization ratio u(v) of v can be calculated by the equation $u(v) = \frac{c(v)}{f(v)}$.

5.2.3 Handling Interactive Query

Mining the appropriate locations to place the billboards requires several rounds of involvements by users (R2). Note that all the input parameters except I_{tv} and I_{vt} are defined by users when they are interacting with the solution generator. More specifically, the model supports three particular interactive queries, namely, (1) pinning the preferred vertices to the final result directly, (2) removing the unsatisfied vertices from the candidates in next round query, and (3) setting the temporal filter to consider the trajectories within that specified period only. The *k*-location and τ-budget constraint query can provide these interfaces through a few minor modifications. For the removal operation, we can modify the input parameter V_{can}. For the pin operation, we can add another parameter V_{pin} to initially allocate its vertices to the final result and update the coverage value of all vertices. The vertices that should be removed or pinned depend on the users who utilize visualization to facilitate their judgments. Similarly, for temporal filter, we can add a parameter T_{period} to filter all trajectories that are not within the specified period.

6 VISUAL DESIGN

This section describes a set of visualization techniques that assist users in generating solutions for billboard placements and comparing them.

6.1 Solution Generator

Users are desired to leverage the computational power of machines to find the optimal solutions (R2). The candidate solutions , however, are exponential in the number of locations, leading to a huge search space for seeking the optimal solutions. Thus, an efficient search and pruning strategy is requested to solve the problem. Moreover, the optimal locations cannot be solely computed by machines, as users typically have their own requirements (R2). Thus, a highly interactive system is needed to allow efficient and smooth interactions with machines. The solution generator (Fig. 1(A, B, C)) combines visualization and mining techniques to enable users to explore the solution space as they desire. It is mainly composed of three sub-views, namely, the dashboard, map, and solution preview view.

6.1.1 Dashboard View

The dashboard view (Fig. 1A) shows the information of the current solution under construction. From top to bottom, the view shows the dataset, target area panel, solution area (the area considered for placing billboard) panel, and the parameter setting panel, which includes the budget (number of billboards for *k*-location query and cost for τ-budget constraint query), normal trajectory weight, target trajectory weight, temporal filter (weekday/weekend), and speed filter.

6.1.2 Map View

Map-centred exploratory approach is very common when it comes to making multiple criteria spatial decisions [20], as it can provide intuitive insights into environment. Thus, we provide a map view (Fig. 1B) that integrates the real roadmap pictures at multiple scales by using Google Maps API; users are allowed to change the map style into satellite or plain map (Fig. 4D) for different purposes. In addition to the roadmap layer, we add three other feature layers, namely, *the area drawing*, *heatmap*, and *marker layer*.

The area drawing layer provides the function of drawing polygons in the map, enabling users to specify the target and solution areas (i.e., the areas considered for placing billboards) in the forms of red and blue polygons, respectively (Fig. 2). The editing and removing interactions in the area are also supported on the dashboard (i.e., target area and solution area panel).

To determine the befitting solution areas (R1), an evident visual representation of spatio-temporal distributions of the target trajectories are required. Thus, we add a heatmap layer, where we provide users with two types of density maps, namely, the OD and road heatmap (Fig. 2). The OD heatmap represents the density of the target trajectories' pick-up and drop-off geolocations through color encoding, with the dense red areas indicating frequent pick-up and drop-off events. The road heatmap shows the density of the target trajectories on each road segment using the same color encoding in the OD heatmap. To facilitate in-depth analysis, both heatmaps can support OD filter (i.e., showing the trajectories to or from the target areas) and temporal filter (i.e., showing the trajectories that occurred in weekdays or weekend) by selecting different options on the top-right corner of the map.

A marker layer presents the locations selected by users or machines with blue markers on the corresponding locations on the map (Fig. 4D). The markers can be removed by clicking them on the dashboard. Users are also allowed to pin their favorite locations on the map. When a location marker is clicked, the detailed information of a billboard location is shown (Fig. 2), including the statistics of passing trajectories, the surrounding POI information, and the OD heatmap of the passing trajectories (R3).

6.1.3 Solution Preview

Users want to know their operating records and the general performance information of each previously generated solutions, so they can explore the solution space efficiently based on previous experience.

The solution preview saves the user's previous settings on each solution and also provides the statistical information for each of them. Fig. 1C illustrates the solution preview, where each box represents one solution. The parameters are shown inside the boxes. A bar chart is horizontally aligned on the top of each box where eight attributes are shown, including the number of billboard (N), cost (C), average speed (S), traffic volume (V), value for money (M), reach (R), OTS (O), and GRP (G). The bar charts are assigned with different categorical colors to indicate which solutions they belong to. When users hover over a bar of one solution, the bars indicating the same attribute values of other solutions would be highlighted to facilitate comparison (as shown in Fig. 6 #1 - #6). The horizontally laid boxes enable users to perform a rough-level comparison among candidate solutions; thus, the solutions with poor performances can be easily detected and deleted (R4). Besides, users are allowed to directly edit any solution by adding or deleting billboard locations from it (R2).

6.2 Solution Explorer

Users have to prepare multiple solutions and then analyze the relative merits of these solutions with customers. This way, customers can select the most satisfying one according to their criteria and preferences. Thus, a visualization tool is necessary for users to conduct in-depth comparative analysis of the candidate solutions. The solution explorer (Fig. 1(D, E, F)) is designed to meet the requirements of R4 to R7. It assists users in conducting multi-perspective comparative analysis among different solutions. The solution explorer comprises three sub-views, namely, the solution, location, and ranking view. These views can be linked to assist users analyze and compare the multiple candidate solutions in the same time at different levels of details.

6.2.1 Solution View

The solution view aims to provide users with a visual summary of each solution and the relationships among them (R5, R6). Meanwhile, this view is a pivot that connects the location and ranking view, thereby enabling users to explore the solutions from multiple perspectives and assisting them determine the optimal solution immediately (R5).

Glyph Design. A glyph design that can reveal the general performance of a solution is required to visually summarize the important features of each solution. However, numerous features may influence the performance of a solution. If we show all of these features simultaneously, then the glyph can be visually complex, which may impose cognitive burden to the users working memory and reduce the task performance. After discussions with our collaborators, we identify the key information that should be visually encoded, namely, reach (weekday and weekend), traffic speed, cost, and POIs. These features enable users to obtain a quick overview of a solution.

A familiar metaphor can greatly enhance comprehension and reduce the cognitive burden on working memory. Inspired by the dashboards of vehicles, we design a novel radial-based [11] visual metaphor to represent a solution, as all traffic-related attributes are naturally related to vehicles. Fig. 3A shows our design. Users often generate at most a dozen solutions and people can efficiently distinguish a dozen colors [37]. Hence, we use the color of the inner circle to indicate the particular solution (consistent with the colors used in the solution preview). The radius of the inner circle represents the total cost by default (the encoding attribute can be specified by users).

A radius heatmap is attached inside the outer circle, which is similar to the speed meter in a vehicle's dashboard; the dark red color indicates high speed. We use a pointer to clearly indicate the average traffic speed for all the trajectories passing the locations of the solution. In addition, the arc area outside the outer circle represents the volume of weekday and weekend reach. The arc is constantly at 180°, the scale is equal to the total number of target trajectories. The arc subareas lying to the left and to the right of vertical dashline represent the weekday and weekend volume of the target trajectories, respectively. The blue area encodes the proposition of the covered target trajectories among all target trajectories. We add a scale to the arc area to help perceive the proposition accurately, and attach nine small radial nodes to the remaining space surrounding the glyph to show nearby POIs in different POI categories (nine categories are selected by our collaborator). We use the size of the nodes to encode the number of POIs. From top to bottom, they are public transport, academic, residence, hotel, sport, life service, shopping, catering, and automobile service.

Design alternatives: Fig. 3 presents several alternative designs that have been evaluated. Design C uses a horizontal ruler on the top to show the reach volume information. The advantage is the use of length to represent the magnitude value, which is perceived more accurately than angle [37]. However, it is not symmetrical and discords with circle. Our collaborators did not like the design. Thus, we proposed two other designs (Design D-E) using arc length to represent the reach volume information. Design D utilizes two scales (one is on the volume arc and the other one lies insider the circle). Although design D cuts down on the use of color, the two radial scales stay too close and can easily confuse users. In contrast, design A and E similarly use a scale on the volume arc and a radial heatmap inside the glyph to represent different scales. Design E uses the number of blocks to visually encode the average traffic speed without using a pointer. However, our collaborators reported that it is not as effective as design A for comparing speeds among multiple solutions.

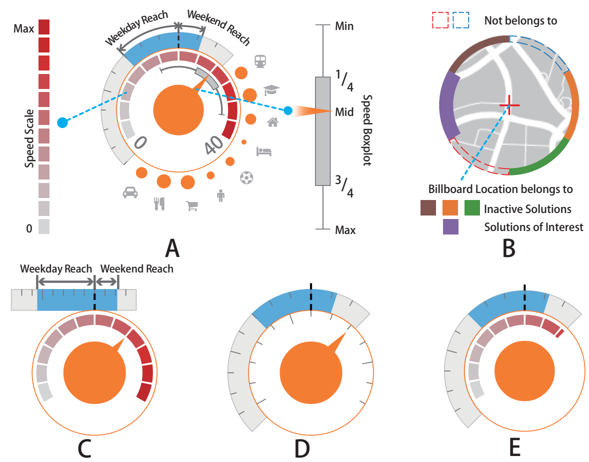

Fig. 3. (A) A dashboard-like glyph design to summarize the key features of each solution. (B) A radial location glyph to present a selected billboard location. (C, D, E) show the alternative solution glyph designs.

Layout: The solution view based on a scatterplot layout allows users to obtain a quick overview of the relationships among solutions (R5). We compute the similarities among different solutions and utilize Multidimensional Scaling (MDS) [27] to create the layout. We also use a well-established method to eliminate the overlap issue [14]. The similarity between two solutions depends on how many the same trajectories are covered. Thus, we can treat the trajectories covered in the two solutions as two sets A and B. We calculate the set similarity by $|A \cap B / min(|A|, |B|)|$ [45]. Two sets are more similar when they share more elements. In the resulting layout, the closer the solution glyphs stay, the more common target trajectories share.

6.2.2 Location View

From the solution view, users can detect the relationships among different solutions with respect to the similarity on the covered target trajectories. Users may want to further analyze the relationships at the fine-grained location level. For example, they may want to know how the locations in each solution distribute geographically, and which solutions the locations of interests belong to. Such information can help users identify the commonalities and differences among the solutions from the perspective of locations (R5). Therefore, we design a visualization that can assist users in this kind of tasks.

To this end, we showed a few designs to our collaborators. Lineset style [35] design (i.e., locations shown in one solution are linked using a line) was first introduced; however, this design has several issues. First, the users are unfamiliar with the map style. Second, this technique suffers from a scalability issue that can lead to visual clutter when a location is shared by many solutions. Moreover, this technique can only show the geographical distribution, whereas our collaborators want to acquire more information, such as the cost of each location. To address these concerns, we had several rounds of discussions with our collaborators. Finally, we designed a new radial location glyph (Fig. 3E) and derived a layout algorithm by extending Dorling cartogram [13] to show the value-by-location maps effectively.

In Dorling cartogram, the relative geographical positions among different objects are considerably preserved. In our case, we use a circle to represent a selected billboard location. By employing the layout of Dorling cartogram, the relative geographical positions of these billboard locations are preserved (Fig. 5B). Fig. 3B shows a location glyph. The primary visual variables should be used to encode the primary data attributes; as suggested by our collaborators, we use the radius of a circle to encode the cost information by default (the encoding attribute can be specified by users). In addition, we fill the circle with a plain-style roadmap in a novel manner. The center of the circle represents the billboard location marked as a red cross. The visual design enables users to immediately identify the road environment around the billboard. To reveal the set relations between solutions and locations, we attach a set of radial color bars surrounding the circle to indicate the solutions that the location belongs to. For the solutions to which users pay attention, the corresponding color bars will be highlighted with increased thickness.

6.2.3 Ranking View

Users need a flexible ranking tool to help them quickly identify good solutions they desire (R7). In particular, all performance-related indicators should be displayed on demand. The system should enable users and customers to freely adjust each attributes weight because they may have different opinions on the performance indicators. The ranking should be updated accordingly and instantly. Moreover, users explained that showing only one value for each attribute is insufficient; they need to analyze further details with respect to locations contained in each solution. For example, users want to know the reach of each location in every solution for the reach indicator. This type of information provides users with detailed insight into how good a solution is regarding a given indicator (R4).

The ranking view visualizes the detailed performance related to the attributes of each solution, including the number of billboard, cost, speed, volume, VFM, reach, OTS, GRP, slowness (inverse of speed for ranking use), in a highly organized and interactive tabular form. To support location-level comparison, we embed boxplots [4] into the matrix. This new design enables users to glean insight into the relative performance of the solutions. This view together with the aforementioned two views (i.e., the location view and the solution view) empower users to quickly find the optimal solutions as they desire (R7).

Fig. 1F shows a matrix-based view, which is inspired by lineup [18]. The first column lists all the candidate solutions with color bars on the left indicating different solutions. The color scheme used is consistent with that in the solution and location views. Other columns display the attribute values, which are normalized and encoded by the length of bars. The width of a column represents the weight users assigned to the attribute. Users can click on the header of a column to rank the solutions by the associated attribute. The columns can also be grouped by right clicking on their headers to rank the solutions by the weighted sum of the attributes in the group. For example, in Fig. 2, the columns reach, OTS, GRP, VFM, Volume, and slowness are grouped; the solutions are ranked by weighting the sum of these attributes. When users are interested in an attribute, they can click on the body of that column , which can expand as a boxplot to show the statistical distributions of the corresponding attributes of the locations in each solution (see the column reach in Fig. 1F). In particular, the minimum, first quartile, median, third quartile, and maximum values are revealed.

6.3 Interactions

User interactions supported by the system are summarized as follows.

Details-on-demand is supported by SmartAdP to facilitate exploration in solution space and analysis of patterns at different levels of details. In the general map view and location view, users are allowed to click any location to see its detailed information. In addition, the boxplot is only shown on demand.

Filtering and highlighting enable users to focus on the information of interests. In SmartAdP, users can filter trajectories using a temporal filter. In particular, the traffic OD information can be further filtered based on whether the trajectories enter or leave for the target areas. The highlighting feature is supported in every view. For example, when hovering on a bar in a solution preview's barchart, the bars encoding the same attribute in other solutions are highlighted. When hovering on a glyph in the solution view, the corresponding locations and solutions are highlighted in the location and ranking views.

Linking connects the three views in the solution explorer. When users click on a glyph in the solution view or a solution color bar in the ranking view, they can be visually linked with the contained locations in the location view. Edge bundling technique [50] is applied to reduce the visual clutter and increase the readability.

7 EVALUATION

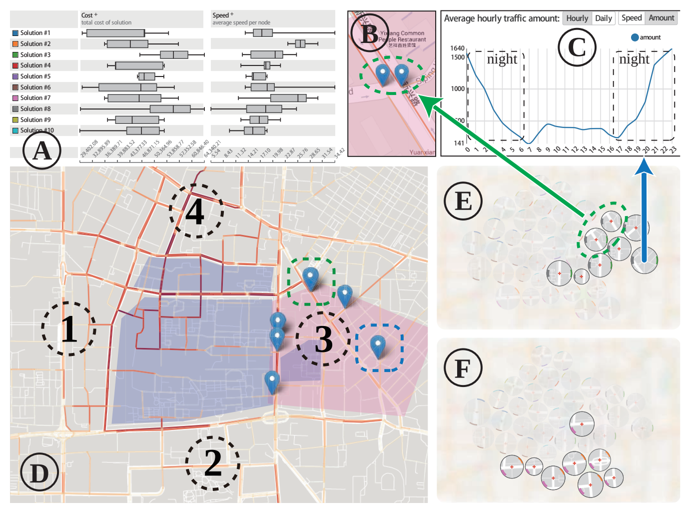

Fig. 4. Three patterns detected. First, the road heatmap (D) highlights the road stretches heavily passed by the target trajectories where areas 1 and 4 have more target trajectories than areas 2 and 3. Second, areas 2 and 3 tend to have higher speed and cost, respectively (A, E, and F). Third, the locations marked with the green and blue rectangles are selected unexpectedly. The green one (B) shows two locations in close proximity and the blue one has a lot of traffics at night (C).

7.1 Case Studies

We conducted the case studies with our domain experts; they were familiar with our designs and databases. As a note, that the cost mentioned in this section denotes monthly cost.

7.1.1 Exploring the solution space

In this scenario, the experts used SmartAdP to explore the solution space. They selected a few target areas (the blue areas in Fig. 4D), in which four famous universities are located.

Exploring the distribution of target trajectories. The experts sought to determine which areas are appropriate for placing billboards (R1). To this end, they first needed to understand the distribution of the target trajectories. From the OD heatmap (Fig. 1B), the experts identified five different areas: (1) an area with a large railway station, (2) a business area that caters to dining or shopping activities, (3) a research center that is home to a number of scientific research institutions, (4) a park area for leisure or exercise, and (5) a central area in which the placement of billboards is expensive. From the road heatmap (Fig. 4D), the experts found that the road stretches in areas 1 and 4 were used more heavily than those in areas 2 and 3.

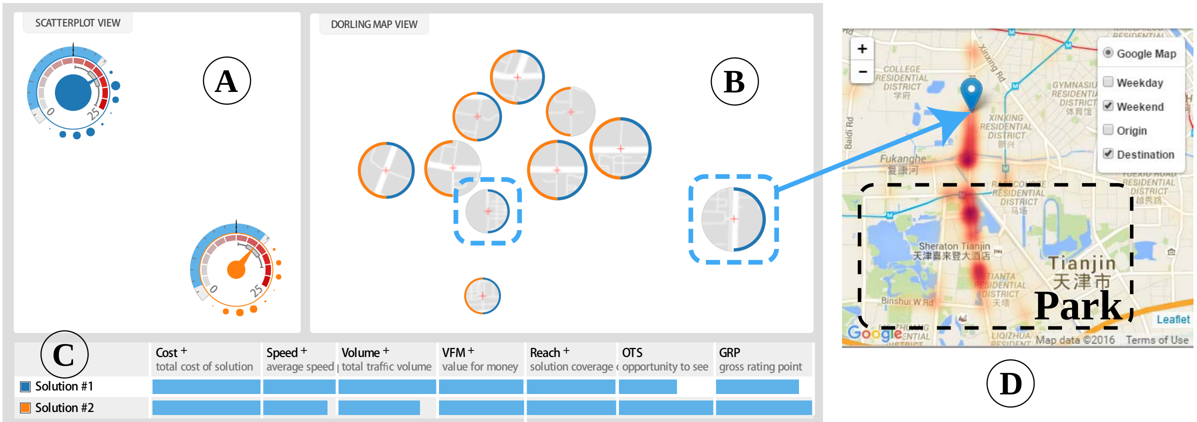

Fig. 5. The two locations marked with the blue rectangles present the difference between the weekend (blue) and weekday (orange) solutions. The location detail view (D) shows that the destinations of the passed trajectories on weekends are mostly parks.

Comparing solution areas for billboard placements. The experts aimed to determine how good each selected solution area was (areas 1-4 in Fig. 4D). A global area depicted as a combination of the four areas was also included. The experts attempted a setting with five billboards and another setting with eight billboards in each solution area (four areas in four directions and one global area). Hence, $5 * 2 = 10$ candidate solutions were formulated, during which process the solution preview provided a quick overview of the performances of the created solutions (R4). To conduct in-depth analyses, the experts switched to the solution explorer, where they identified four clusters (C1-C4 in Fig. 1D) the solution view (R5, R6).

As observed by the experts, the solutions in C3 and C4 clearly have a significantly low reach because of the short blue arcs. Moreover, the solutions in C3 achieve fast speeds (see the speed pointer and radial boxplot inside the associated glyphs). Fast speeds are not desired by the experts because fast speeds do not allow sufficient time for viewing billboards. The speeds are reflected in the boxplot in the ranking view (Fig. 4A), where the medians of solutions #2 and #7 are much higher than others. The location view (with the circle size encoding the speed in Fig. 4F) reveals that the locations in C3 are in the bottom area. The solutions in C4 entail a relatively high cost (the glyphs have larger inner circles in Fig. 1D), as the associated locations in the location view (with circle size encoding cost by default in Fig. 4E) are close to the city central area (area 5 in Fig. 1B). This can also be validated by the size of POI circles around the glyphs in the solution view (Fig. 1D); typically the larger the POI circles (more POIs around), the more expensive the areas for placing billboards. C1 contains four solutions with higher reach than those in the other clusters because the blue radial arcs are longer. Furthermore, the experts found that the billboards of the solutions in C1 are located near the top of the map (e.g., the billboards of the purple solution in Fig. 1E). Meanwhile, the experts changed the radius of the circle in the location view to encode the reach, and found the circles on the top tend to have larger size than others, which indicates the higher reach. This is in accordance with the OD and road heatmaps. By further checking the boxplot in ranking view (the reach column in Fig. 1F), the experts found that solutions #4, #5, #9, and #10 are of high reach. Solutions #4 and #9 are formulated with the solution area 4 (Fig. 4D). Solutions #5 and #10 are formulated with the global solution area, which also include area 4. This indicates most of optimal locations are within the area 4 in Fig. 4D. C2 contains two solutions whose solution area is area 2 in Fig. 4D. These two solutions perform normally. In summary, areas 4 (Fig. 4D) is preferable solution area for billboard placements.

Improving and editing solutions. When the experts evaluated area 3 in Fig. 4D, they found two unexpected locations selected by the model (R3). First, the two selected locations are too close (marked with the green rectangle in Fig. 4(D, E)). It was explained that the road is made up of several lanes in opposite directions and both directions have large volume of target trajectories. Whereas, the experts thought that placing two billboards in close proximity is not necessary. Second, despite the relative large traffic volume of the location (marked with the blue rectangle in Fig. 4(D, E)), it is mainly attributed to the traffics at night (Fig. 4C. Experts zoomed in on the map, thereby finding the location surrounded by a number of bars. As the illumination at night is not as good as daytime, daytime traffic is preferable. Thus, the experts removed the two unexpected locations.

7.1.2 Finding the optimal solution

This case is aimed to demonstrate the effectiveness of the solution explorer in identifying optimal solutions (R7). Note that the target areas in this case is the same as the previous one.

Weekday vs. Weekend. Driven by the previous experience, the experts quickly generated two solutions. The two solutions cover the same solution areas, that is, the global area combining area 1-4 Fig. 4D. The two solutions differ in terms of the setting of the temporal filter. One solution only considers weekday trajectories, whereas the other solution only considers weekend trajectories. As indicated by the glyphs in Fig. 5A and the performance indicators in Fig. 5C, these two solutions (blue for weekend and orange for weekday) perform similarly because all the indicators, except OTS, are quite similar. By further checking the location view (with circle size encoding cost in Fig. 5B), the experts found that the difference between the two solutions was mainly caused by the two locations highlighted in the blue rectangle. They were particularly interested in the one on the right side, as it was expensive (large circle) and far from the commonly selected locations. Hence, the experts explored the OD heatmap of the location (Fig. 5D), and found that the destinations of the trajectories that passed the location on weekends were mostly parks. The experts inferred that people tended to visit parks on weekends. This result also explains why the location was picked by the model (R3), that is, the location showed a large number of target trajectories on weekends. People going out on weekends, especially those visiting parks, tend to be in a good mood and are highly likely to be influenced by advertising content. All the experts preferred solution #1 despite solution #2 being slightly better in terms of the OTS and GRP.

Dispersed Strategy vs. Clustered Strategy. The experts formulated a series of candidate solutions given a budget of $\$380,000$. Two advertising campaign strategies were employed: (1) a dispersed strategy, which involves placing billboards in a large region to increase the reach as much as possible; (2) a cluster strategy, which involves placing billboards in a small region to increase the OTS as much as possible. Using SmartAdP , the experts generated three candidate solutions for each strategy with different parameter settings (Fig. 6). Specifically, solutions #1 to #3 were generated by adopting the dispersed strategy (the solution area is nearby area 4 in Fig. 4D). Solution #1 and #2 were generated with a weight ratio (the target trajectory weight to the normal trajectory weight) of 5 and 100, respectively. Solution #3 involved the addition of a speed filter ($\leq 15km/h$). Solutions #4 to #6 were generated with the same settings while limited in a small region in area 4 of Fig. 4D.

Through the solution explorer, the experts identified an outlier (the blue solution) in the solution view (C2 in Fig. 2 Solution View). For the locations in the blue solution, they are distributed dispersedly in the location view (Fig. 2 Location View) and show high traffic volume (Fig. 2 Ranking View), as well as a low reach (the glyph in solution view has short outer radial arc). The other 5 solutions form two clusters, where the solutions with the same advertising strategy cluster together (C2 and C3 in Fig. 2 Solution View). From the glyphs, the experts knew the speeds of the brown and green solutions were slow owing to the speed filter, but their reach was relatively low in comparison with that of the other solutions. The experts further used the ranking function in the ranking view to determine the optimal solution. The results created by ranking the solutions according to a single column (attribute) are listed in Fig. 6. Generally, the dispersed strategy (#1 - #3) performs much better than the clustered strategy (#4 - #6) in terms of reach, whereas the clustered strategy achieves a higher OTS than the dispersed strategy. However, the solutions have different rankings when ranked by different attributes. Hence, determining the optimal solution at first glance was difficult and as such the experts further ranked the solutions with a group of specified columns (attributes). The view allowed the experts to adjust the weight of each column flexibly in the group. Finally, they obtained a ranking of the solutions (Fig. 2 Ranking View), where the GRP and VFM are of higher weights than the other attributes. As a result, solution #5 generated with clustered strategy was selected.

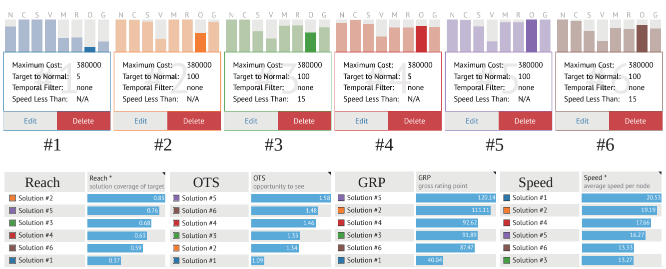

Fig. 6. The solution preview (top) displays the parameters and statistics of each solution. The bottom shows the solution ranks with the different attributes selected.

7.2 Domain Expert Interview

We collected the feedback of the domain experts by conducting one-on-one interviews. The feedback was summarized as follows.

Visual Design and Interactions. The experts confirmed that our system is well designed and user friendly. In particular, our metaphor-based glyph design received high praises from the domain experts. They believed that the system could be easily understood by users with different backgrounds. EA commended, "Today's consumers are not so suggestible for the sake of an increasingly accessible Internet. Your visualizations can be an incredible tool to clearly demonstrate the upside of our service." EC praised the smooth interactions, commenting "SmartAdP integrates many advanced navigation and interaction techniques, enabling me to smoothly work with machines."

Usability and Improvements The experts appreciated our system and found the functions provided by SmartAdP quite useful. They all agreed that our system is useful in not only making adjustments to campaigns but also interpreting the data and leveraging the findings to benefit their clients and brands. Apart from the aforementioned, our experts also provided some valuable suggestions. EA mentioned, "In many times, I found several flaws of candidate solutions in the solution explorer. It would be nice if I can edit solutions directly in this interface." EB further suggested an advanced solution merging function would be helpful, as two solutions are sometimes complementary.

8 DISCUSSION

Our evaluation demonstrates the effectiveness of SmartAdP . Nevertheless, there is still space for improvement.

The system does not support visual comparison of hundreds of solutions, because we use colors to differentiate candidate solutions and people cannot effectively distinguish over one dozen colors [37]. Nevertheless, we believe that our design works for most cases because users typically do not generate many solutions. Likewise, the location view suffers from the same problem with the increase of billboards. A possible solution to the problem is to employ a hierarchical strategy to handle numerous circles. Additionally, the current system mainly focuses on comparison at the solution level and does not support displaying values of multiple criteria for locations directly. Although the location view and embed boxplots in the rank view can facilitate comparison among locations, they can only display one attribute at the same time. Therefore, we plan to investigate more design choices for showing multiple attributes of locations in the rank view.

Travel direction should be considered when a road is so wide that billboards are visible only from one side of the road. The current system is capable of dealing with that situation for the reason that the passing-by trajectories of the locations on the two opposite sides of a wide road are counted separately. More specifically, for a wide road in the road network data, it is always treated as two road segments with opposite directions; hence, the locations on the two road segments count only the trajectories from one direction. However, travel direction can also influence the manner in which we place a billboard in a given location. For instance, we may also need to take into account several additional factors, such as the billboard orientation and the height off the ground. These factors are not considered currently but are worthy of being explored in the future.

Regarding the model itself, the greedy heuristic is suggested to extract the locations. Nevertheless, our work does not limit the utilization of other more sophisticated methods to further improve the quality of the results. For example, the anytime refinement [51] is one of the possible methods. More advanced methods enabling users to interactively train and improve models are also worth further investigations. For taxi data, despite its prominent advantages, it cannot adequately represent all the mobility movement patterns. This issue is common in prior studies that utilize taxi trajectories. Hence, we are interested in integrating additional types of data, such as location-based social network data or the information about regional functions. This requires more efforts on solving the problem of heterogeneous data fusion [47].

Although our work is primarily designed for the billboard location selection problem, it can be easily adapted for other similiar problems, such as selecting locations of retail stores or restaurants. In these similiar problems, a number of good locations can be recommended by adapting our novel visualization-driven model with the data that can reveal human mobility patterns, such as taxi trajectories. Our visualization approach, including solution generator and solution explorer, can also be tailored and extended to support visual editing and comparison of different sets of locations.

9 CONCLUSION

In this paper, we systematically study the problem of identifying the optimal billboard locations using massive trajectory data. Closely working with the end users enables us to derive two major challenges facing billboard advertising planners, namely, creating and comparing multiple solutions in an immediate and accurate manner. Hence, we present SmartAdP, an interactive visual analytics system that combines a new application-driven mining model with several well-designed visualization and interaction techniques. We conduct case studies and expert interviews to demonstrate the system. Positive feedback and in-depth insights show the usefulness and effectiveness of our system.

Acknowledgments

The work is supported by National 973 Program of China (2015CB352503), the Fundamental Research Funds for Central Universities (2016QNA5014), National Natural Science Foundation of China (No. 61502416), the research fund of the Ministry of Education of China (188170-170160502), 100 Talents Program of Zhejiang University, RGC GRF16241916, ITS/170/15FP, and a grant from Microsoft Research Asia.

References

[1] APN Outdoor. `http://goo.gl/Aln5AB`.

[2] Lamar. `http://goo.gl/P86jkQ`.

[3] MongoDB: 2dsphere Indexes. `https://goo.gl/wlarPf`.

[4] S. Few. Distribution displays, conventional and potential, 2014. Technical report. `http://goo.gl/zy8uAo`.

[5] N. Andrienko and G. Andrienko. Informed spatial decisions through coordinated views. *Information Visualization*, 2(4):270–285, 2003.

[6] N. Andrienko and G. Andrienko. Spatial generalization and aggregation of massive movement data. *IEEE TVCG*, 17(2):205–219, 2011.

[7] N. Andrienko and G. Andrienko. Visual analytics of movement: An overview of methods, tools and procedures. *Information Visualization*, 12(1):3–24, 2013.

[8] W. Chen, F. Guo, and F.-Y. Wang. A survey of traffic data visualization. *IEEE TITS*, 16(6):2970–2984, 2015.

[9] Z. Chen, H. T. Shen, X. Zhou, and J. X. Yu. Monitoring path nearest neighbor in road networks. In *Proc. ACM SIGMOD*, pages 591–602, 2009.

[10] C. Collins and S. Carpendale. Vislink: Revealing relationships amongst visualizations. *IEEE TVCG*, 13(6):1192–1199, 2007.

[11] W. Cui, H. Qu, H. Zhou, W. Zhang, and S. Skiena. Watch the story unfold with textwheel: Visualization of large-scale news streams. *ACM TIST*, 3(2):20, 2012.

[12] K. Deng, K. Xie, K. Zheng, and X. Zhou. Trajectory indexing and retrieval. In *Computing with Spatial Trajectories*, pages 35–60. Springer, 2011.

[13] D. Dorling. Area cartograms: their use and creation. In *The Map Reader: Theories of Mapping Practice and Cartographic Representation*, pages 252–260. Wiley, 2011.

[14] T. Dwyer, K. Marriott, and P. J. Stuckey. Fast node overlap removal. In *Proc. Graph Drawing*, pages 153–164, 2005.

[15] U. Feige. A threshold of ln n for approximating set cover. *Journal of the ACM (JACM)*, 45(4):634–652, 1998.

[16] N. Ferreira, J. Poco, H. T. Vo, J. Freire, and C. T. Silva. Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips. *IEEE TVCG*, 19(12):2149–2158, 2013.

[17] M. Gleicher, D. Albers, R. Walker, I. Jusufi, C. D. Hansen, and J. C. Roberts. Visual comparison for information visualization. *Information Visualization*, 10(4):289–309, 2011.

[18] S. Gratzl, A. Lex, N. Gehlenborg, H. Pfister, and M. Streit. LineUp: Visual analysis of multi-attribute rankings. *IEEE TVCG*, 19(12):2277–2286, 2013.

[19] X. Huang, Y. Zhao, J. Yang, C. Zhang, C. Ma, and X. Ye. TrajGraph: A graph-based visual analytics approach to studying urban network centralities using taxi trajectory data. *IEEE TVCG*, 22(1):160–169, 2016.

[20] P. Jankowski, N. Andrienko, and G. Andrienko. Map-centred exploratory approach to multiple criteria spatial decision making. *International Journal of Geographical Information Science*, 15(2):101–127, 2001.

[21] H. Jeung, Q. Liu, H. T. Shen, and X. Zhou. A hybrid prediction model for moving objects. In *Proc. ICDE*, pages 70–79, 2008.

[22] D. Karamshuk, A. Noulas, S. Scellato, V. Nicosia, and C. Mascolo. Geo-spotting: Mining online location-based services for optimal retail store placement. In *Proc. ACM SIGKDD*, pages 793–801, 2013.

[23] H. Katz. *The media handbook: A complete guide to advertising media selection, planning, research, and buying*. Routledge, 2014.

[24] J. Kehrer and H. Hauser. Visualization and visual analysis of multifaceted scientific data: A survey. *IEEE TVCG*, 19(3):495–513, 2013.

[25] J. Kehrer, H. Piringer, W. Berger, and M. E. Gröller. A model for structure-based comparison of many categories in small-multiple displays. *IEEE TVCG*, 19(12):2287–2296, 2013.

[26] S. Khuller, A. Moss, and J. S. Naor. The budgeted maximum coverage problem. *Information Processing Letters*, 70(1):39–45, 1999.

[27] J. B. Kruskal. Nonmetric multidimensional scaling: a numerical method. *Psychometrika*, 29(2):115–129, 1964.

[28] J.-G. Lee, J. Han, X. Li, and H. Gonzalez. TraClass: Trajectory classification using hierarchical region-based and trajectory-based clustering. In *Proc. VLDB*, pages 70–79, 2008.

[29] J.-G. Lee, J. Han, and K.-Y. Whang. Trajectory clustering: A partition-and-group framework. In *Proc. ACM SIGMOD*, pages 593–604, 2007.

[30] H. Liu, Y. Gao, L. Lu, S. Liu, H. Qu, and L. M. Ni. Visual analysis of route diversity. In *Proc. IEEE VAST*, pages 171–180, 2011.

[31] S. Liu, W. Cui, Y. Wu, and M. Liu. A survey on information visualization: recent advances and challenges. *The Visual Computer*, 30(12):1373–1393, 2014.

[32] Y. Lou, C. Zhang, Y. Zheng, X. Xie, W. Wang, and Y. Huang. Map-matching for low-sampling-rate GPS trajectories. In *Proc. ACM SIGSPATIAL*, pages 352–361, 2009.

[33] J. Malczewski. *GIS and multicriteria decision analysis*. Wiley, 1999.

[34] J. Malczewski. Gis-based multicriteria decision analysis: a survey of the literature. *International Journal of Geographical Information Science*, 20(7):703–726, 2006.

[35] W. Meulemans, N. H. Riche, B. Speckmann, B. Alper, and T. Dwyer. KelpFusion: A hybrid set visualization technique. *IEEE TVCG*, 19(11):1846–1858, 2013.

[36] T. Munzner. A nested model for visualization design and validation. *IEEE TVCG*, 15(6):921–928, 2009.

[37] T. Munzner. *Visualization Analysis and Design*. CRC Press, 2014.

[38] R. Scheepens, C. Hurter, H. Van De Wetering, and J. J. Van Wijk. Visualization, selection, and analysis of traffic flows. *IEEE TVCG*, 22(1):379–388, 2016.

[39] R. Scheepens, N. Willems, H. Van de Wetering, G. Andrienko, N. Andrienko, and J. J. Van Wijk. Composite density maps for multivariate trajectories. *IEEE TVCG*, 17(12):2518–2527, 2011.

[40] C. Shi, Y. Wu, S. Liu, H. Zhou, and H. Qu. LoyalTracker: Visualizing loyalty dynamics in search engines. *IEEE TVCG*, 20(12):1733–1742, 2014.

[41] C. Tominski, H. Schumann, G. Andrienko, and N. Andrienko. Stacking-based visualization of trajectory attribute data. *IEEE TVCG*, 18(12):2565–2574, 2012.

[42] Y. Wang, Y. Zheng, and Y. Xue. Travel time estimation of a path using sparse trajectories. In *Proc. ACM SIGKDD*, pages 25–34, 2014.

[43] Z. Wang, M. Lu, X. Yuan, J. Zhang, and H. Van De Wetering. Visual traffic jam analysis based on trajectory data. *IEEE TVCG*, 19(12):2159–2168, 2013.

[44] Y. Wu, F. Wei, S. Liu, N. Au, W. Cui, H. Zhou, and H. Qu. OpinionSeer: interactive visualization of hotel customer feedback. *IEEE TVCG*, 16(6):1109–1118, 2010.

[45] M. A. Yalcin, N. Elmqvist, and B. B. Bederson. AggreSet: Rich and scalable set exploration using visualizations of element aggregations. *IEEE TVCG*, 22(1):688–697, 2016.

[46] K. Zheng, Y. Zheng, N. J. Yuan, and S. Shang. On discovery of gathering patterns from trajectories. In *Proc. IEEE ICDE*, pages 242–253, 2013.

[47] Y. Zheng. Methodologies for cross-domain data fusion: An overview. *IEEE transactions on big data*, 1(1):16–34, 2015.

[48] Y. Zheng. Trajectory data mining: an overview. *ACM TIST*, 6(3):29, 2015.

[49] Y. Zheng, L. Capra, O. Wolfson, and H. Yang. Urban computing: concepts, methodologies, and applications. *ACM TIST*, 5(3):38, 2014.

[50] H. Zhou, P. Xu, X. Yuan, and H. Qu. Edge bundling in information visualization. *Tsinghua Science and Technology*, 18(2):145–156, 2013.

[51] S. Zilberstein. Using anytime algorithms in intelligent systems. *AI magazine*, 17(3):73, 1996.