BRPVis: Visual Analytics of Bus Route Planning based on Perception of Passenger Travel Demand

*

Fig. 1. The user interface of BRPVis: (a) Panel view supports inputting the OD positions and adjust weight quantification results. (b) Projection view helps planner discover routes of interest in the optimal route set. (c) Route overview allows planners to perform a stop-level overview comparison of all routes. (d) Ranking view displays the ranking result of routes and enables planners to compare and analyze routes from multiple perspectives. (e) Map view presents the hotspot areas of passenger travel and detailed information about each route and stop. (f) Comparison view allows planners to compare the original property values of multiple routes of interest.

Abstract—The priority development of public transportation is an effective way to mitigate environmental pollution and improve the operational efficiency of urban transportation. However, determining the optimal bus scheme for route planning remains challenging due to the difficulties in achieving a reasonable placement of candidate bus stops and the large solution space of feasible bus routes. Moreover, the candidate routes produced by automated methods requires planners to conduct comprehensive evaluation and comparison with considering multiple influencing factors, which is a complex process. These issues motivate us to propose a visual analytics framework for bus route planning. Designing such a framework poses three major challenges: a) ensuring the rationality and accuracy of candidate bus stop mining, b) achieving equilibrium optimal bus route set solution with multi-objective mutual constraints, c) interactive exploratory analysis yields the optimal route planning scheme. For challenge a), we propose a candidate bus stop mining method that perceives passenger travel demands and consider multiple factors to achieve accurate stop placement. For challenge b), we construct a multi-objective optimization model for bus route planning to obtain the optimal planning schemes under different goals. For challenge c), we design a visual analytics system BRPVis and develop interactive feedback mechanisms in the progressive exploration process to help planners understand the candidate route planning schemes and explore the influencing factors. We evaluate BRPVis with the real city traffic datasets through three case studies to demonstrate the effectiveness in the task of bus route planning.

Index Terms—Bus Route Planning; Attribute weight quantification; Visual Analytics; Smart City

1 Introduction

Public transportation plays an important role in solving environmental pollution and traffic congestion. Due to the continuous changes of human travel patterns, the bus system need to be improved iteratively. Many cities have opened dedicated buses and customized buses, such as commuter buses. Bus route planning is required to consider the combined effects of multiple factors comprehensively, and seek a balance among the needs of travelers and operating costs. However, determining a rational bus route remains a difficult task, due to the following issues:

I1. Accuracy of travel demand perceiving and bus stop mining needs to be improved. Traditional route designing approaches rely on

XXX, XXX(corresponding author), XXX, XXX and XXX , XXX, XXX, China. E-mail: XX

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

census or travel intention surveys, where the relatively subjective and small samples cannot fully meet the travel needs of passengers. With the widespread use of Global Positioning System (GPS) technology in urban taxis, many research works mined travel demand [8] and movement patterns [2] based on large-scale taxi GPS data, which is more objective and effective for bus route planning. However, the existing methods of prerceiving travel demand and then deploying bus stops neglected the overall distribution of bus stops and their service scopes, whose accuracy still needs to be improved.

I2. It is required to ensure the efficiency and accuracy of generating bus route candidates from a huge solution space with considering multiple objectives. Based on the candidate bus stops, bus route planning generates a huge solution space, i.e., the optimal set of routes needs to be sought among a large number of route solutions. Moreover, multiple objectives such as oute running time and passenger flow, should be considered for obtaining usable bus route candidates. Therefore, how to ensure the efficiency and accuracy of solving the optimal set of routes is one of the main challenges in bus route planning.

I3. An exploratory environment is required to select the satisfied routes with analyzing multiple influencing factors quantitatively. Quantitative analysis of route influencing factors, such as the cumulative passenger flow of the routes and the number of stops and other attributes, is extremely important to compare the differences between candidate routes and to select reasonable route planning solutions. Moreover, to meet the requirements of planning departments, planners need to understand and adjust the influencing factor weights by combining their experience, and conduct comprehensive evaluation and comparison of routes. To the best of our knowledge, this requirement has been not well supported by existing works of bus route planning.

To solve the above problems, we propose a visual analytics framework for bus route planning with perceived passenger travel demand, which improves the accuracy of bus stop mining and route formulating, quantifies and analyzes the influence of route attribute weights on the development of route planning schemes, and provides a valuable decision basis for planners to carry out bus route planning. For issue I1, we analyze passenger daily travel data from a data-driven perspective and propose a grid clustering-based passenger travel demand perception method to improve the accuracy of candidate bus stop mining. For issue I2, a multi-objective model for route planning is constructed with multiple objectives, and the key idea of Non-dominated Sorting Genetic Algorithm-II (NSGA-II) multi-objective optimization is introduced into the problem of solving the optimal set of bus routes. For issue3, Ranking SVM [7] is introduced to quantify the route attribute weights, and we design a interactive feedback mechanism inspired by Podium [19]. It can help planners improve the accuracy and credibility of route ranking recommendations by changing the weights and conducting subjective ranking operations. Through integrating the above methods and results, we design and develop BRPVis, a visual analytics system to support multi-level comparison of bus routes and assist route planners in exploring optimal bus route planning solutions. In addition, case studies and evaluations based on real-world datasets demonstrate the effectiveness of the system. The main contributions of this paper are as follows:

- We proposed a candidate bus stop mining method based on GPS data to perceive passenger travel demand, which can take into account the connectivity of passenger travel hotspots and the service scope of bus stops, enhancing the rationality and accuracy of candidate bus stop mining.

- We proposed a multi-objective optimization method for bus route planning, to obtain the optimal planning schemes under different goals, improving the accuracy of the generated routes and the efficiency of the optimal set solution, with respect to the traditional PBS method.

- We designed a visual analytics system BRPVis and developed interactive feedback mechanisms in the progressive exploration process to help planners understand the candidate route planning schemes and explore the influencing factors.

2 Related Work

2.1 Bus Route Design

Bus route design is a classical research problem in the field of transportation and urban planning [6, 28], which aims to make the designed bus routes better meet passenger travel demand with limited resource allocation and budget under certain constraints, such as minimizing the length of operating routes while maximizing the cumulative passenger flow of the routes [15]. In the past decades, bus route design was mainly based on manual survey methods to calculate passenger travel demand and station traffic [3], weighted summation methods were used to consider operating costs and passenger travel costs. At present, most of the related research work mainly focuses on data-driven methods to mine public travel demand and achieve efficient planning of bus routes based on big traffic data [1].

Data-driven bus route planning is usually divided into two steps: candidate bus stop mining and route generation. First, most research works determine the passenger travel demand in the planning area through traffic data and mobility data [17], and mine a number of candidate bus stop locations from the travel big data through clustering methods such as DBSCAN [27] and CFSFDP [12]. For example, Xia et al [26] proposed SP-DBSCAN, a clustering method based on the improvement of DBSCAN, and the generated clustered regions can provide help for passengers to choose the location of the bus ride and better meet their travel demand. Next, based on the mined bus stops, route solution are generated usually with the OD specified by planners [9, 16, 22]. For example, Chen et al [1] used a random search method to make the generated bus routes balanced in terms of both running time and cumulative passenger flow when planners specify the origin and destination. Fan et al [29] described the bus route design problem using a nonlinear mixed integer mathematical model and a genetic algorithm. However, the complexity of bus route planning and the fact that the selection scheme of bus routes is influenced by a combination of subjective (e.g., the total number of route stations) and objective (e.g., the degree of emphasis people attach to an attribute) factors, which make it difficult for the optimal route planning scheme derived from the above method to meet the needs of the route planning department. This paper will introduce a route attribute weight quantification method based on existing research to achieve the above objectives and obtain a practical and optimal bus route planning solution.

2.2 GPS Trajectory Data Mining and Visualization

Research work based on taxi trajectory data has become one of the hot research topics in the transportation field. Wang et al [20] summarized the latest research trends in trajectory data, including trajectory preprocessing, storage, common trajectory analysis tools, and trajectory clustering, and explored closely related data analysis tasks to illustrate the significance of research work on trajectory data mining.

Taxi trajectory data mining has been widely used in travel pattern discovery, route recommendation, and urban planning, etc. Wu et al [25] mined the influencing factors affecting taxi detour behavior from taxi GPS trajectory data, proposed a map matching-based detour clustering method to process the trajectory data, and established a multi-layer road analysis system, aiming to illustrate the changes in the spatial and temporal distribution of taxi detour characteristics and perform statistical analysis. Wang et al [21] proposed an interactive GPS trajectory-based visual analysis system for urban traffic congestion, developed a perceptual strategy for extracting and deriving traffic congestion information, provided multiple visualization views for intuitively exploring and analyzing the traffic conditions of an entire large city, and presented the main spatial and temporal correlated. Ferreira et al [5] proposed a new model to support users to visualize taxi trips from a data-driven perspective, in addition, the model supports origin-to-destination queries so that the movement of people in the whole city can be studied. In the work of Liu et al [13], a novel interactive visual analytic framework based on data mining models using large-scale GPS trajectory data was proposed with the aim of solving the billboard siting problem. Based on the taxi trajectory data, this study fully exploits the potential information of taxi trajectory data by considering the acceptable walking

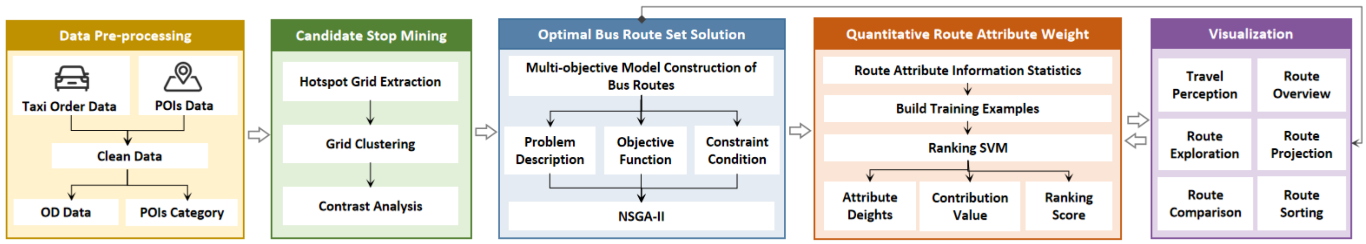

Fig. 2. Analysis pipeline of the system includes data pre-processing, candidate bus stop mining, optimal route set solving, route attribute weight quantification and visual analystic.

range of passengers (i.e., the service area of bus stops), the connectivity of stops, the accumulated passenger flow, and the type and number of POIDs in the reachable area, so as to generate more reasonable route planning solutions.

2.3 Multi-Attribute Ranking Visualization

Data ranking visual analytics is a hot research in the field of visualization. For example, Lee et al [10] designed an interactive visual analytics system for analysis tasks such as traffic congestion detection, monitoring, and prediction, which can rank roads based on road speed, estimated travel time, road flow through a table view to provide decision support for planners. In visual analytic tasks based on ranking, there are related research works that have designed various visualization systems as well as views specifically for ranking [4, 11, 23, 24]. Podium [19] enables users to rank subsets of data through their own holistic understanding of multi-attribute data items through a table view as well as a control panel, thus using the system to rank the subsets and train a ranking model that reflects the user's subjective preferences of the sorting model, and thus analyze the user's sorting behavior. In a recent study, Shrestha et al [18] designed FairFuse, a visualization system for generating, analyzing, and validating rankings, which encodes relevant metrics through a novel visualization design as a way to achieve ranking visualizations that users can intuitively understand and support interactive exploration of ranking results. Liu et al [14] proposed a visual analysis system RankAxis, which systematically combines ranking and projection methods to facilitate mutual interpretation of these two techniques and supports multi-attribute data exploration, based on real-world case studies, expert feedback, and user studies that demonstrate the effectiveness of the system. The ranking visualization design in this paper is mainly applied to multi-attribute ranking of bus routes. Based on previous research work, the ranking work in this paper allows users to achieve subjective ranking of subsets of data items and understand their subjective ranking behavior through model training, i.e., to obtain quantitative results of attribute weights for each ranking condition as a representation of users' subjective preferences, and combines novel visual view presentations to facilitate intuitive comparison and support interactive exploration and analysis by users.

3 Overview

3.1 Datasets

The dataset used in this paper includes urban taxi order data and urban POI data. **Taxi order data** of one city in the time span from August 1st, 2021 to August 26th, 2021, including boarding and alighting latitude and longitude, boarding and alighting time, and order number. A total of 2,251,274 valid data, with 4,502,548 boarding and alighting behaviors. **POI statistics data** of one city in 2021, including location latitude and longitude, address name and category code. 10 categories are catering services, accommodation, scenic spots, education and culture, institutions and organizations, living services, malls and supermarkets, leisure and entertainment, companies and enterprises, and transportation facilities, with 71,480 pieces of data.

3.2 Analytical Tasks

In the past six months, we worked closely with three domain experts (EA-EC). EA is an analyst from the transportation department that provides the experimental data. EB is a researcher of public transportation planning, who has extensive experience in defining the constraints and objectives of bus routes. EC is an urban computing expert, whose interest is traffic analysis. After literature review, we discussed with experts by weekly video meetings to understand the requirements and iteratively refine the analysis tasks, which were summarized as:

T1: Reveal hotspots for passenger daily travel. An visual overview of the hotspot areas for city taxi pickups and drop-offs is needed to support planners in perceiving the passenger travel demand.

T2: Explore the valuable bus stop. Candidate bus stops should be extracted by considering factors such as service area, connectivity, type and number of accessible POIs, and the accumulated passenger traffic. At the same time, planners also need to inspect the detailed information of the stops.

T3: Solve the optimal bus routes under multiple objectives and support the route overview. The system should support in generating and overall displaying the optimal set of route planning solutions with specified origins and destinations under multiple objectives, including maximize the cumulative passenger and minimize the travel time.

T4: Support ranking analysis of candidate routes. The system should be able to generate route rankings based on quantitative attribute weights of route planning, support the operation of subjective ranking and recommend more routes that meet the focused needs through ranking.

T5: Support comparative analysis between route solutions. The system should support multi-perspective comparisons between different candidate routes from the whole to the details, with interactive explorations.

3.3 System Analysis Pipeline

Based on the above analytis tasks, we designed and implemented a visual analysis system. The analysis pipeline of this system is shown in Fig. 2, including five modules: data pre-processing, candidate bus stop mining, optimal route set solving, route attribute weight quantification and visualization. In the **data pre-processing module**, the above data sets are pre-processed to obtain passenger travel OD data and POI data. In the **candidate stop mining module**, a grid clustering method considering multiple factors is proposed to mine candidate bus stops. In the **optimal route set ssolution module**, a multi-objective model for route planning is constructed, and the optimal route set solving method based on NSGA-II is proposed. In **quantitative route attribute weight module**, statistical route attribute information is used to obtain attribute weight quantification and ranking results through the Ranking SVM method. In the **visualization module**, system supports planners to interactively explore candidate route solutions through a multi-map linkage mechanism. It also supports changing the route attribute weights to provide personalized decision support for route planners.

4 Bus Route Planning and Attribute Weight Quantification Method

4.1 Candidate bus stops mining

Based on urban taxi order data, this paper implements candidate bus stop mining through two steps. which are (1) passenger travel hotspot grid extraction; (2) candidate bus stops generation based on grid clustering.

4.1.1 Passenger Travel Hotspot Grid Extraction

In this paper, we proposed a method to aggregate the pick-up and drop-off points into a grid and mark the connectivity degree (CD) and pick-up and drop-off records (PDR) of each grid to determine the traffic hotspot areas. Firstly, city datum latitude and longitude Lat and Lon are obtained, which indicate the latitude and longitude of the top left vertex of the smallest rectangle containing the city geography space, respectively. The grid $(long_g, lat_g)$ to which each boarding point (lon_i, lat_i) belongs is calculated by Equation 1:

$$\begin{cases} long_g = floor((lon_i - Lon)/step) * step + Lon \\ lat_g = floor((lat_i - Lat)/step) * step + Lat \end{cases} \tag{1}$$

where $step = 0.0001° \approx 11m$, grid set G is obtained by traversing all the pick-up and drop-off points, and the number of pick-ups and drop-offs in this process is counted as the grid pick-up and drop-offs PDR. At the same time, it is known that there are at most eight grids around any grid, therefore,the number of grids around each grid is marked as the grid connectivity CD, which value range is [0,8]. Finally, for the data in $Sect$3.1, a total of 293,496 on or off hotspot grids are obtained. That is, $|G| = 293496$, as shown in Fig.3a.

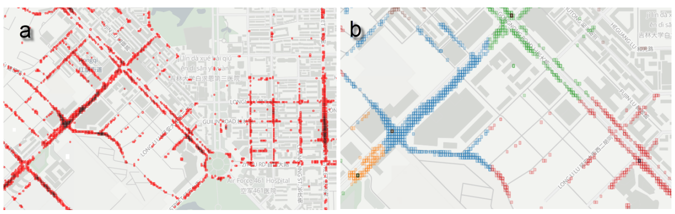

Fig. 3. (a) Grid of boarding and alighting hotspots, the darker color means more boarding and alighting occurs in the grid. (b) Grid clustering results, whose color black indicates the location of candidate bus stops, and the color indicates their service area.

4.1.2 Candidate Bus Stop Generation

We proposed **Identification of Candidate Bus Stops (ICBS)** method based on grid clustering to improve the rationality and accuracy of candidate bus stop mining, as shown in Algorithm 1, where $MaxRadius$ is the service radius of each station, which will be analyzed in Sect 4.3 for parameter sensitivity. Besides, it make the candidate bus stops selected at locations with intensive boarding and alighting behavior and high passenger flow by setting $w_1 = w_2 = 0.5$. Finally, 2,357 candidate bus stops were obtained within the city, as shown in Fig. 3b.

According to the China Urban Road Traffic Planning and Design Specification, the service area of buses is recommended to be within the radius of 500 to 800 meters. In this paper, to verify the effect of different values of $MaxRadius$ on the clustering results, 20m steps were taken for the experiment. As shown in Fig. 4, all four evaluation indicators show that the clustering effect is best when 500m is taken. Considering that the acceptable walking range for passengers is as small as possible, this paper sets $MaxRadius = 500m$.

Finally, for each candidate bus stop in the set C, we counte the number of passenger transfers between stops based on order data stored in the matrix $Passenger$, and the bus travel time between stops stored in the matrix $Traveltime$, which is set to be 1.5 times longer than the taxi travel time.

4.2 Bus Route Planning and Attribute Weight Quantification Method

4.2.1 Problem Statement

Based on the above method, we get the set of candidate bus stops $S = \{s_1, s_2, s_3, ...\}$, where $s_i = (lon_i, lat_i)$; when the user specifies the bus origin s_o and destination $s_d, where s_o, s_d \in S$; the bus routes generated

---

Algorithm 1 Identification of Candidate Bus Stops (ICBS)

Input: $G = \langle g_1, g_2, ..., g_n \rangle, g_i = (lon_i, lat_i, CD_i, POR_i)$, n is the total number of grids.
Output: Candidate bus stop set $C = \langle c_1, c_2, ..., c_n \rangle$

set $MaxRadius = 500m, C = \emptyset$;
Add g_1 to C
for $i = 0$; $i < n$; $i++$ **do**
  for $j = 1$; $j < len(c)$; $j++$ **do**
    Calculate the distance $SD(g_i, c_j)$ between the grid g_i and the cluster center c_j,find the cluster center grid with the smallest distance c;
    if $SD(g_i, c) < MaxRadius$ **then**
      $g_i \in c$;
    else
      Add g_i to C;
    end if
    Calculate the values of all k grids in the cluster to which the cluster center grid c belongs, and obtain the index with the maximum value according to the following equation:
$$new = arg\max_h \left[w_1 * \frac{CD_h}{8} + w_2 * \frac{PDR_h}{\sum_{h=1}^{k} PDR_h} \right]$$
    Add g_{new} to C and delete c;
  end for
end for

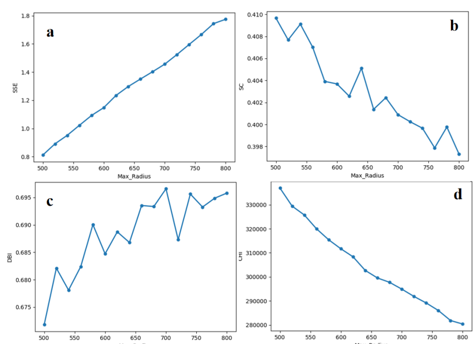

Fig. 4. Variation curve of each indicator with $MaxRadius$, (a)SSE, (b)SC, (c)CHI, (D)DBI.

between these two stops are represented as $R = (r_1, r_2, ..., r_n) \in P_{od}$, where $r_1 = o, r_n = d, s_{r_1}, s_{r_2}, ..., s_{r_n} \in S$, all feasible routes form the set P_{od}, and the research problem in this paper is to make P_{od} close to the optimal solution by model solving.

4.2.2 Multi-Objective Model for Bus Route Planning

Objective function construction. The first objective is to minimize the travel time, the bus travel time ΔT_R between two stops is shown in Equation 2, and the meaning of each symbol in the equation is shown in Table 1.Besides, T can be obtained from $Traveltime$ matrix. $R = (r_1, r_2, ..., r_k)$ indicates the set of stations that have been added to the route and not reached the destination s_d, s_n is the new stations added to the route. Therefore, the first objective function is to minimize the route running time, as shown in Equation 3.

$$\Delta T_R = h + T\left(s_{r_{|R|}}, s_n\right) \tag{2}$$

$$\Delta \min T_R = h * (|R| - 2) + \sum_{i=1}^{|R|-1} T\left(s_{r_i}, s_{r_{i+1}}\right) \tag{3}$$

The second objective is to maximize the cumulative passenger flow of the route, and for each new bus stop, the cumulative increase in passenger flow is shown in Equation 4. we consider the cumulative passenger flow, not just the transfer of passenger flow between two adjacent stops. F can be obtained from the $Passenger$ matrix. Therefore,

Table 1. Notation Description

Notation	Description
$P_{od} = \{R_1, R_2, ..., R_n\}$	Optimal route set (Pareto set).
$R = (r_1, r_2, ..., r_n)$	One bus route.
$s_i = (lon_i, lat_i)$	Candidate bus stop location.
$S = \{s_1, s_2, ...\}$	Set of candidate bus stops.
$T = (s_i, s_j)$	Travel time between the s_i and s_j.
h	Parking waiting time.
$F = (s_i, s_j)$	The passenger flow between the s_i and s_j.

the second objective function is to maximize the cumulative passenger flow of the route, as shown in Equation 5.

$$\Delta D_R = \sum_{i=1}^{|R|} F\left(s_{r_i}, s_n\right) \tag{4}$$

$$\Delta \max D_R = \sum_{i=1}^{|R|-1} \sum_{j=i+1}^{|R|} F\left(s_{r_i}, s_{r_j}\right) \tag{5}$$

Constraints. We follow the approach outlined in the prior study [1]to build five intuitive criteria to determine the feasibility of the transit route construction:

1) s_i must be at most δ meters away from s_j;

2) s_j must be farther from the origin s_o and closer to the destination s_d than s_i along the direction of $s_o \rightarrow s_d$;

3) s_j must be farther from the origin s_o than s_i;

4) s_j must be closer to the destination s_d than s_i;

5) s_i should be closer to s_j than any other station before s_i to avoid zigzags.

We set $\delta = 3km$ and adopt the road distances to generate the station graphs instead of the Euclidean distances to emulate a more realistic setting.

4.2.3 NSGA-II Based Optimal Route Set Solving

The key steps of NSGA-II genetic algorithm are divided into three main parts, which are fast non-dominated sorting, individual congestion distance calculation, and elite retention strategy, as shown in Fig.5. Firstly, population graph (in this paper refers to bus routes) is initialized and more routes are generated by selection, crossover and mutation operations. Next, the elite retention strategy retains the good individuals into the children, which is achieved by fast non-dominated sorting and individual crowding distance calculation, and the operation of population hierarchy is realized by calculating the objective function value of individuals, which finally makes the individuals retained in the population closer and closer to the optimal set of solutions. When the similarity of the optimal set is located at 1 and constant, or after satisfying the number of iterations, the set of children at this time is the optimal route set.

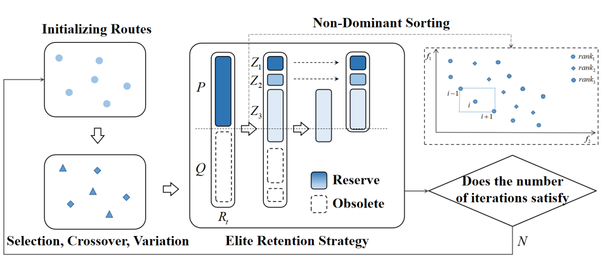

Fig. 5. Workflow of NSGA-II based bus route planning model solution method.

In this paper, we use the key idea of NSGA-II to solve the multi-objective optimization model for bus route planning.This method is described in Algorithm 2.

Algorithm 2 Optimal Route Set Solving Method

Input: $S = \{s_1, s_2, s_3 ...\}, s_i = (lon_i, lat_i)$;
Passenger Matrix,Traveltime Matrix;
Origin s_o and destination $s_d, s_o, s_d \in S$.

Output: Optimal bus route set P_{od}.

Set the total number of initialized customized bus routes is N, iteration number $g = 1$,and maximum iterations g_{max}.
Set $P_{od} = \emptyset$.
for $c = 0; c < N; c{+}{+}$ **do**
  $selectNode = 0$;
  while $selectNode \neq d$ **do**
    Set route $R = (r_1)$, $r_1 = o$;
    $P(s_i^* | \langle s_1, s_2, \ldots, s_j \rangle) = \dfrac{\sum_{m=1}^{j} F(s_k, s_i^*)}{\sum_{i=1}^{|S^*|} \sum_{k=1}^{j} F(s_k, s_i^*)}$;
    $allowSize = length(P)$;
    $selectP = Random(0, 1)$;
    $sumP = 0$;
    $startIndex = Random(allowSize) + 1$;
    while $sumP < selectP$ **do**
      $sumP = sumP + P[(startIndex - 1)\%allowedSize]$;
      $startIndex = startIndex + 1$;
    end while
    $selectNode = (startIndex - 2)\%allowedSize$;
    Add $selectNode$ to R;
  end while
  Add R to P_{od};
  $c = c + 1$;
end for
while $g < g_{max}$ **do**
  Calculate the value of each route R_i in the set P_{od} according to Equation 3 and Equation 5;
  Crossover : for each route, a station is randomly selected to exchange with another route while satisfying the constraints, resulting in a new route;
  Mutation : for each route, two stations are randomly selected in one of the route stations, replacing the original route with a new route between these two stations and satisfying the constraints to obtain a new route;
  Calculate the value of new routes according to Equation 3 and Equation 5;
  Update the set P_{od} according to NSGA-II;
  $g = g + 1$;
end while

4.2.4 Route attribute weight quantification method

Based on Ranking SVM [19], we implemented weight quantification around five attributes of the generated bus routes, including cumulative route passenger flow (attr1), route travel time (arrt2), route length (attr3), number of stops (attr4), and number of accessible POIs (attr5). Firstly, the optimal set of routes is obtained while the five attribute values of each route are counted in this paper, as shown in Fig.6a. Then, the initialized ranking by the attribute of cumulative passenger flow.

Training example constructed by Equation 6 is shown in Fig. 6b, where $d_i - d_j$ indicates the attribute difference vector of any two routes, and 1 or -1 indicates that the current ranking of d_i is higher or lower than that of d_j. Inputting this data into Ranking SVM, we can obtain the results of each attribute weight $w_1, w_2, ..., w_5$ after recalculating the ranking score $r(d_i)$ of each route by Equation 7 as the cumulative sum of the products of attribute values and corresponding weights, and obtaining the ranking results according to their magnitudes.

$$\left(d_i - d_j, \theta\right), \theta = \begin{cases} 1 r(d_i) \geq r(d_j) \\ -1 r(d_i) < r(d_j) \end{cases} \tag{6}$$

$$r(d_i) = w * d_i = \sum_{j=1}^{m} w_j * a_j \tag{7}$$

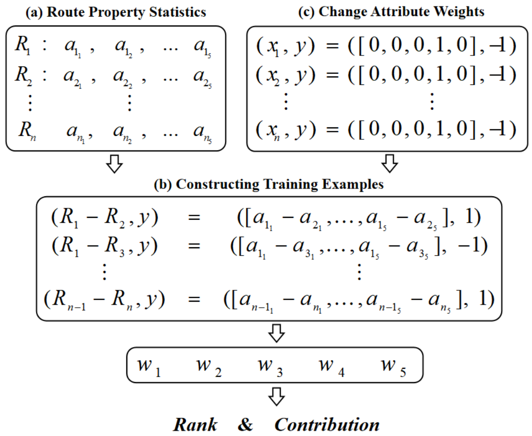

Fig. 6. Schematic diagram of Ranking SVM applied to bus route attributes.

Contribution value of a data item helps the user to understand how much of the ranking score of that data item is due to a given attribute. Equation 8 represents the normalized contribution of an attribute a_j to the ranking score of a data item d_i. Where l is the index of the attribute with the largest attribute score for the data item.

$$\hat{C}\left(d_{ij}\right) = \frac{\left|d_{ij}w_j\right|}{max_l\left\{\left|d_{ij}w_l\right|\right\}} \tag{8}$$

We also provides an interface to change the attribute weights, which is achieved by adding a new training tuple to the training instance, as shown in Fig.6c. Where the difference at the position corresponding to the changed attribute is marked as 1 and the rest of the positions are 0. If the user wants to emphasize the attribute, the training tuple $(-x, y = -1)$ is constructed, instead, construct the training tuple $(-x, y = 1)$. The symbols for this technique are chosen from the Ranking SVM update algorithm and are empirically validated to ensure that they can produce meaningful changes to the weight vector.

4.2.5 Results and Evaluation

In order to verify the convergence of the model solving method, this paper introduces the concept of set similarity. If the optimal set of routes is located at 1 and remains unchanged after a certain number of executions, the algorithm is said to converge as shown in Equation 9, where A and B denote the optimal set of routes before and after the update.

$$Similarity\,(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{9}$$

5 Visual Design

BRPVis is developed to support task proposed in Sect.3.2, which consists of six parts,including a map view (Fig.1e) is used to show the passenger travel heat grid, route connections, and route details. Route projection view (Fig.1b) encodes the route attributes to help planners quickly find the route of interest. Route ranking view (Fig.1d) is used to show the total ranking score of each route, attribute score and contribution of each attribute to the route ranking. Routes overview (Fig.1b) allows to display all route connections. Route comparison view (Fig.1f) helps planners to compare the routes they interested in terms of both the functionality of the area where it can be reached and route attributes. Configuration panel (Fig.1a) supports planners input OD location latitude and longitude as well as the display and adjustment of route attribute weights.

5.1 Projection View

Projection view of the optimal route set is designed as shown in Fig.1b, with the horizontal axis indicating the total route running time and the vertical axis indicating the cumulative passenger flow, each glyph in the figure represents an optimal route, and the radius of the center circle is coded as the number of other routes ruling the collection, i.e., how many routes are better than the other routes in the collection in two objectives, and the larger the radius of the circle, the more routes are ruled. The three arcs represent the attribute values of "number of stations", "route length" and "number of accessible POIs", which are coded by the arc length, with the longer radius representing the larger attribute value. In addition, in order to highlight the "rank1" routes, the routes are colored, while the rest of the rank routes are not colored,which Support interactive view of route id and click to select route to view details.

5.2 Overview of Route

We design a overview of route to facilitate planners to compare each route at the station level, as shown in Fig.1c.Each row represents a candidate route, and each square represents a station passing through each route. The side length and color of the square encode the cumulative traffic of the station, and the longer side length and darker blue color of each square represent more cumulative traffic of the station. The real geospatial distance between stations is mapped to the length of the connecting route between two stations and the shade of the color, the longer each connecting route the darker the orange color represents the real distance. At the same time, the bus travel time between any two stations is coded as the thickness, and the thicker the route, the longer the travel time between the two stations. It can help users to compare the differences between the routes and provide help for further comparison and analysis visually.

5.3 Ranking View

Based on the quantification of route attribute weights and the ranking results, this paper designs a ranking view, as shown in Fig.1d. Each row represents an optimal route, and the first column shows the total ranking score of the route, which is coded by the length of the orange horizontal bar graph, with the longer length indicating the higher total ranking score, i.e., the higher ranking; the second to the sixth columns show the five attributes of the route, which are coded by the length of the blue horizontal bar graph, with the longer length representing the higher score on each attribute, while taking the form of alternating dark blue and light blue to avoid visual confusion. In addition, the contribution value is coded as the distance between the black vertical route and the leftmost end of each attribute bar. It can upport for clicking on the route name to select the route of interest for detailed display in other views. The user can also click on the route name to drag and drop to change the ranking of the route, and the system will construct training data based on the current ranking and retrain the model to get the new weight results and ranking results.

5.4 Map View

Map view contains of two forms, such as the thermal grid view shown in Fig. 1, where the grid color codes the number of passenger boardings and alightings within the grid, with darker blue representing more boardings and alightings occurring at that location, and the location of candidate bus stops coded in red. Planners can use this view to perceive passenger travel demand and support click selection of origins and destinations. The route detail view shows the route's station connections, as shown in Fig 1. In addition, it supports viewing the details of each station, showing the type and number of accessible POIs at each station through a pie chart, when clicking on each station through a route graph showing the distribution of passenger flow at the station and showing the average cumulative passenger flow at the station and the travel time to the station in the information box.

5.5 Comparison View

User can click in other views to select the route of interest to add to the comparison view. It consists of two parts, the bar chart shown in Fig.1f is used to visualize the number of reachable categories of POI for each route. It can help the analyst understand the overall functionality of each route and also visually compare the differences between them. At the same time, the raw attribute values of the routes are presented through bubble charts, as shown in Fig.1g, where one row represents a route and five columns represent each of the five attributes of the route, whose values are encoded as the size of a circle radius, the larger radius represents larger attribute values, and they are distinguished by color.

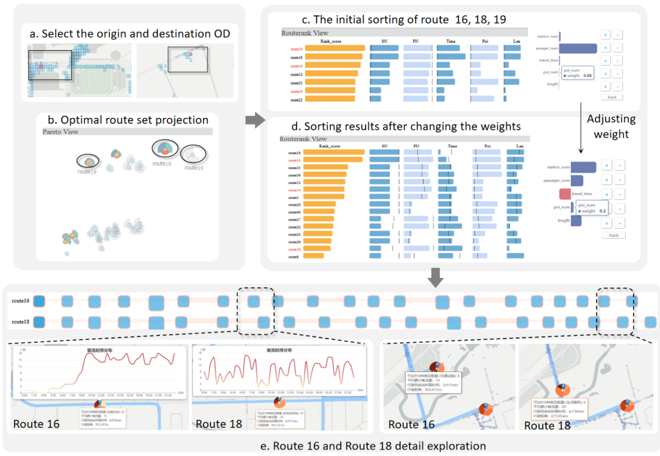

Fig. 7. Case1: Exploration of Bus Route Planning for Custom OD.

5.6 Panel View

Panel view includes input boxes for latitude and longitude of the origin and destination for performing route planning, as shown in Fig.1a. It supports the planner to select OD in the hot grid map. The lower part of the panel also visualizes the results of the quantification of the weights of the five route attributes. Positive weights are coded in blue and displayed on the right side, negative weights are coded in red and displayed on the left side, and the length of the bars encodes the magnitude of the weight values. In addition, planners can adjust the attribute weight results by using the "+" and "-" buttons on the right side of each attribute, and choose to increase or decrease the corresponding attribute weights, then click the "Rank" button to retrain the model to get the new weights and ranking results.

6 Case Studies

6.1 Exploration of Bus Route Planning for Custom OD

In this case study, EB was invited to explore the passenger travel area with map view, then he wanted to select the train station as the starting point for route planning because there have the higher passenger flow as we all konw. Meanwhile, he found that there have a candidate bus stop near a square within the geographical area of universities. Considering the possibility of students traveling back to school as well as activities, he selected it as the destination, as shown in Fig.7a. Next, he looked at projection view, as shown in Fig.7b. There are three routes that he interested, in which route 16 and 19 are rank1 route in the Pareto optimal route set and not dominated by other routes, while route 19 has a very high ruling score and the value of each attribute of route 16 is significantly better than other routes. Although route 18 is not the rank1 routeit has height cumulative traffic. Next, he looked at the current ranking of each route, which is in the first, second and sixth positions, as shown in Fig.7c. The attribute that contributes most to their ranking is "PN". Therefore, he reduced their weight and increased the weight of POI, as shown in Fig.7. The weight of POI attribute increases from 0.069 to 0.2, and the weight of PN attribute decreases significantly. In addition, other attribute weights also change, such as the travel time weight becomes negative, i.e., it has a negative impact on the route ranking. Looking at the updated ranking results, as shown in Fig.7d, we can found that the ranking of route 19 dropped significantly to the 14th place, mainly due to its larger "Time" value. Route 16 and 18 remain in the top two positions, with the same number of accessible POIs and cumulative passenger traffic, and route 18 is in first place thanks to its smaller running time.

Next, planner observered route overview which shown that they had same station number, while route 18 takes more time to travel to the 8th station and has a thicker route, as shown in Fig.7e. Further, as shown in Fig.7e, at the eighth station, route 16 has an advantage in terms of the type of POI accessible (accommodation residential P2), average cumulative passenger flow 15, and time required. By clicking on the route graph of the temporal change of the station traffic, it could be seen that the number of passengers is significantly more regular and stayed above 12 most of the time compared to route 18. At the twentieth station, route 16 also has the same advantage and would pass through the station of route 18, so the planner identified route 16 as the bus route to be determined under this OD, and this solution can provide them with decision support.

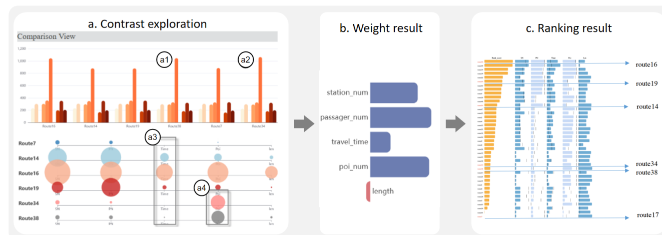

Fig. 8. Case Study2: Subjective Ranking Behavior Analysis.

6.2 Subjective Ranking Behavior Analysis

This case study was conducted by EB from the opposite perspective of previous data ranking efforts, by inverting the route attribute weights through the analysts' ranking of the routes. On the one hand, it will help he to understand the attributes valued by the subjective ranking performed; on the other hand, for ranking under some criteria, he can find important attributes and further can achieve global ranking according to the current focus. In the previous case, six "rank 1" routes were obtained, and the weighted ranking results were 16, 14, 19, 38, 7, and 34. He can subjectively rank them from another perspective. First, he clicked to select the above routes to add to the comparison view, as shown in Fig.8a. It can be seen that the type of POI reachable between the ODs is mainly "life services", among which, except for route 16, the number of such POIs reachable by routes 34 and 38 is higher, and the total number of POIs reachable is also higher, therefore, he decided to increase the ranking of both routes. In addition, route 34 has a lower running time than route 38, and route 19 has a time advantage over route 14, so the user defines the ranking after exploration as 16, 19, 14, 34, 38, 7.

Next, the system constructs model training data based on the ranking of these six routes and obtains the results of each attribute weight as shown in Fig.8. First, the highest weights are accumulated passenger flow and reachable POI, which indicates that users' sorting behavior mainly focuses on these two route attributes. Second, looking at the two attributes of travel time and route length that have a correlation, here the user is more concerned about the shorter route length, which is an attribute with negative influence.

Based on this weighting result, the user clicked the "Rank" button and applied it to all routes to get the overall route ranking, and the result is shown in Fig.8. The relative ranking of six routes was the same as the initial ranking of the planner, which means that the ranking results obtained by the training model completely follow the subjective ranking behavior of the user. Route 16 still remained in the first place and is the more dominant route. Under the subjective ranking behavior, the user can explore and analyze these routes in the system in further details based on the results obtained from two different perspectives to get the final satisfactory results.

6.3 Comparative Analysis

In this case study, EA was invited to input the same OD as the actual route in the system, explore the optimal route planning scheme and

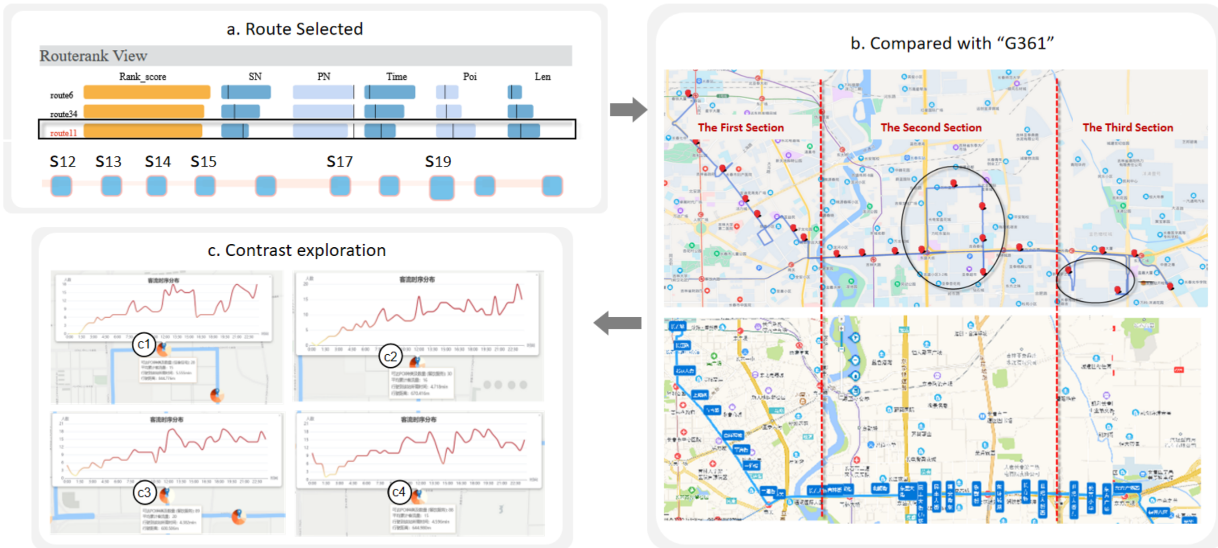

Fig. 9. Case Study3: Comparative analysis with real working route.

then compare and analyze with it to further verify the effectiveness of the system in this paper. Planner seleted a nighttime bus route ”G361”, with a total length of 11.5km and 28 stops, starting from the railway station and ending at Guanghua College Station. After exploring the system interactively, planner selected ”route11” as the nighttime bus route to be determined between this OD. As shown in Fig.9a the route has a lower number of stops (SN) and a higher reachable POI number, while its cumulative passenger volume (PN) is similar to the previous two routes, so it was chosen as the final route to be determined.

The planner can view the route through the real map and compared it with the route ”G361” for analysis. As shown in Fig.9b, the route can be divided into three sections: the first section has basically the same direction and passing places, the second section has a detour through other stations, which is different from the actual route going straight ahead, the third section also has two detours. Therefore, the user analyzed these detours with an overview view of the stops. As shown in Fig.9a , it was found that the detour passes through the stations: s12, s13, s14, s15, s17 and s19. It could be seen through the connecting lines that the driving time of these stations is also less (the connecting lines are thinner), and the cumulative passenger flow of s19 accounts for a much higher percentage than other stations (the side length is longer), so the planner initially perceives that these stations are of some value.Next, the planner mainly looked at the four stations in the second section, as shown in Fig.9. The stations with more POIs were ”Food Service” and ”Residential”. The cumulative traffic volume of each station was also at a high level, with values ranging from 15 to 20 passengers. The route graph shows that the number of passenger trips at each station tends to increase after 21:00 at night, and the number of passengers increases from less to more in the late evening to early morning. Therefore, the planner believe that the existence of detours in the second section is meaningful, which can accumulate more passenger traffic. At the same time, it was observed that s19 could reach more POIs of the type ”company enterprise, these two types of POI for the night bus is more advantageous, respectively involved in the evening dinner and commuting from work and other activities.

In summary, the routes selected by the planners through the system have a high similarity to the planning of the actual route trunk road, and have certain advantages over the actual routes in terms of accumulated passenger flow. It can provide valuable reference basis and decision-making help for route planners when planning actual routes.

7 Discussion

Generalizability. The methods in this paper can be well applied to other GPS trajectory data to perform cluster analysis to solve related problems. In addition, the method of solving the multi-objective model and the attribute weight quantification method in this paper are also applicable to other multi-objective research problems and multidimensional data ranking problems.

Scalability. This paper proposes a visual analytics framework for urban bus route planning that can perception passenger travel demand with certain scalability. Starting from trajectory points for grid partitioning, the grid-based clustering method achieves the purpose of mining candidate bus stops in a larger city-wide and a large amount of trajectory data. The system also integrates the Ranking SVM method to combine multiple attributes of routes for interactive exploration by ranking, which can achieve attribute weight quantification for arbitrary multidimensional data. Each view in the system is adaptively set to complete the presentation of different numbers of data items.

Limitation. For the task of urban bus route planning, the amount considered in this paper is not comprehensive enough in terms of route attributes, and more route attributes can be mined through trajectory data, such as for example, calculating the average speed of each section of the journey in each candidate route scheme through trajectory data, and can grasp the historical congestion level of the road, etc. In addition, although the system can achieve the site level of detail exploration, but it can not support the user to achieve the designation or replacement of sites and other operations, there are certain limitations.

8 Conclusion

This paper proposes BRPVis, a visual analysis system for bus route planning based on large-scale urban traffic data, which supports comparative analysis of bus route options from overview to details and provides decision support for bus route planners. First, we proposes ICBS, a candidate bus stop mining method based on GPS data sensing passenger travel demand. Second, we construct a multi-objective model for bus route planning and generates optimal bus route sets by NSGA-II multi-objective optimization method. On this basic, we introduces Ranking SVM, a route attribute weight quantification method to support quantitative analysis while deeply integrating field experience, aiming to improve the accuracy and credibility of optimal route recommendations and credibility. Further, experiments are conducted using real urban traffic data, and the effectiveness of the method and system in urban bus route planning problems is verified through several case studies and user evaluations.

Acknowledgments

This research was supported in part by National Natural Science Foundation of China under Grant 42171450 and Key R&D Project of Science and Technology Development Plan of Jilin Province under Grant 20210201074GX.

References

[1] C. Chen, D. Zhang, N. Li, and Z. Zhou. B-planner: Planning bidirectional night bus routes using large-scale taxi GPS traces. *IEEE Trans. Intell. Transp. Syst.*, 15(4):1451–1465, 2014. doi: 10.1109/TITS.2014.2298892

[2] E. Chen, W. Zhang, Z. Ye, and M. Yang. Unraveling latent transfer patterns between metro and bus from large-scale smart card data. *IEEE Trans. Intell. Transp. Syst.*, 23(4):3351–3365, 2022. doi: 10.1109/TITS.2020.3035719

[3] T. A. Chua. The planning of urban bus routes and frequencies: A survey. *Transportation*, 12(2):147–172, 1984.

[4] Z. Deng, D. Weng, J. Chen, R. Liu, Z. Wang, J. Bao, Y. Zheng, and Y. Wu. Airvis: Visual analytics of air pollution propagation. *IEEE Trans. Vis. Comput. Graph.*, 26(1):800–810, 2020. doi: 10.1109/TVCG.2019.2934670

[5] 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 Trans. Vis. Comput. Graph.*, 19(12):2149–2158, 2013. doi: 10.1109/TVCG.2013.226

[6] V. Guihaire and J.-K. Hao. Transit network design and scheduling: A global review. *Transportation Research Part A: Policy and Practice*, 42(10):1251–1273, 2008.

[7] R. Herbrich. Large margin rank boundaries for ordinal regression. *Advances in large margin classifiers*, pp. 115–132, 2000.

[8] P. Kothari and A. Alahi. Safety-compliant generative adversarial networks for human trajectory forecasting. *IEEE Trans. Intell. Transp. Syst.*, 24(4):4251–4261, 2023. doi: 10.1109/TITS.2022.3233906

[9] Y. Lai, F. Yang, G. Meng, and W. Lu. Data-driven flexible vehicle scheduling and route optimization. *IEEE Trans. Intell. Transp. Syst.*, 23(12):23099–23113, 2022. doi: 10.1109/TITS.2022.3204644

[10] C. Lee, Y. Kim, S. Jin, D. Kim, R. Maciejewski, D. S. Ebert, and S. Ko. A visual analytics system for exploring, monitoring, and forecasting road traffic congestion. *IEEE Trans. Vis. Comput. Graph.*, 26(11):3133–3146, 2020. doi: 10.1109/TVCG.2019.2922597

[11] Q. Li, K. S. Njotoprawiro, H. Haleem, Q. Chen, C. Yi, and X. Ma. Embeddingvis: A visual analytics approach to comparative network embedding inspection. In R. Chang, H. Qu, and T. Schreck, eds., *13th IEEE Conference on Visual Analytics Science and Technology, IEEE VAST 2018, Berlin, Germany, October 21-26, 2018*, pp. 48–59. IEEE, 2018. doi: 10.1109/VAST.2018.8802454

[12] Z. Li, J. Wang, Z. Shi, and Y. Zou. Relieve the congestion by shuttle bus in rush hours using aggregation clustering algorithm on group travel pattern. *Concurr. Comput. Pract. Exp.*, 31(9), 2019. doi: 10.1002/cpe.4847

[13] D. Liu, D. Weng, Y. Li, J. Bao, Y. Zheng, H. Qu, and Y. Wu. Smartadp: Visual analytics of large-scale taxi trajectories for selecting billboard locations. *IEEE Trans. Vis. Comput. Graph.*, 23(1):1–10, 2017. doi: 10.1109/TVCG.2016.2598432

[14] Q. Liu, Y. Ren, Z. Zhu, D. Li, X. Ma, and Q. Li. Rankaxis: Towards a systematic combination of projection and ranking in multi-attribute data exploration. *IEEE Trans. Vis. Comput. Graph.*, 29(1):701–711, 2023. doi: 10.1109/TVCG.2022.3209463

[15] Q. Q. Liu, Q. Li, C. F. Tang, H. Lin, X. Ma, and T. Chen. A visual analytics approach to scheduling customized shuttle buses via perceiving passengers' travel demands. In *31st IEEE Visualization Conference, IEEE VIS 2020 - Short Papers, Virtual Event, USA, October 25-30, 2020*, pp. 76–80. IEEE, 2020. doi: 10.1109/VIS47514.2020.00022

[16] Z. Ning, S. Sun, M. Zhou, X. Hu, X. Wang, L. Guo, B. Hu, and Y. Kwok. Online scheduling and route planning for shared buses in urban traffic networks. *IEEE Trans. Intell. Transp. Syst.*, 23(4):3430–3444, 2022. doi: 10.1109/TITS.2020.3036396

[17] F. Pinelli, R. Nair, F. Calabrese, M. Berlingerio, G. D. Lorenzo, and M. L. Sbodio. Data-driven transit network design from mobile phone trajectories. *IEEE Trans. Intell. Transp. Syst.*, 17(6):1724–1733, 2016. doi: 10.1109/TITS.2015.2496783

[18] H. Shrestha, K. Cachel, M. Alkhathlan, E. A. Rundensteiner, and L. Harrison. Fairfuse: Interactive visual support for fair consensus ranking. *CoRR*, abs/2207.07765, 2022. doi: 10.48550/arXiv.2207.07765

[19] E. Wall, S. Das, R. Chawla, B. Kalidindi, E. T. Brown, and A. Endert. Podium: Ranking data using mixed-initiative visual analytics. *IEEE Trans. Vis. Comput. Graph.*, 24(1):288–297, 2018. doi: 10.1109/TVCG.2017.2745078

[20] S. Wang, Z. Bao, J. S. Culpepper, and G. Cong. A survey on trajectory data management, analytics, and learning. *ACM Comput. Surv.*, 54(2):39:1–39:36, 2022. doi: 10.1145/3440207

[21] Z. Wang, M. Lu, X. Yuan, J. Zhang, and H. van de Wetering. Visual traffic jam analysis based on trajectory data. *IEEE Trans. Vis. Comput. Graph.*, 19(12):2159–2168, 2013. doi: 10.1109/TVCG.2013.228

[22] D. Weng, R. Chen, J. Zhang, J. Bao, Y. Zheng, and Y. Wu. Pareto-optimal transit route planning with multi-objective monte-carlo tree search. *IEEE Trans. Intell. Transp. Syst.*, 22(2):1185–1195, 2021. doi: 10.1109/TITS.2020.2964012

[23] Y. Wu, Z. Chen, G. Sun, X. Xie, N. Cao, S. Liu, and W. Cui. Streamexplorer: A multi-stage system for visually exploring events in social streams. *IEEE Trans. Vis. Comput. Graph.*, 24(10):2758–2772, 2018. doi: 10.1109/TVCG.2017.2764459

[24] Y. Wu, X. Xie, J. Wang, D. Deng, H. Liang, H. Zhang, S. Cheng, and W. Chen. Forvizor: Visualizing spatio-temporal team formations in soccer. *IEEE Trans. Vis. Comput. Graph.*, 25(1):65–75, 2019. doi: 10.1109/TVCG.2018.2865041

[25] Z. Wu, Y. Li, X. Wang, J. Su, L. Yang, Y. Nie, and Y. Wang. Mining factors affecting taxi detour behavior from GPS traces at directional road segment level. *IEEE Trans. Intell. Transp. Syst.*, 23(7):8013–8023, 2022. doi: 10.1109/TITS.2021.3074976

[26] D. Xia, Y. Bai, Y. Zheng, Y. Hu, Y. Li, and H. Li. A parallel SP-DBSCAN algorithm on spark for waiting spot recommendation. *Multim. Tools Appl.*, 81(3):4015–4038, 2022. doi: 10.1007/s11042-021-11639-9

[27] L. Xiao, X. Fan, H. Mao, C. Xu, P. Lu, and S. Luo. When taxi meets bus: Night bus stop planning over large-scale traffic data. In *7th International Conference on Cloud Computing and Big Data, CCBD 2016, Macau, China, November 16-18, 2016*, pp. 19–24. IEEE Computer Society, 2016. doi: 10.1109/CCBD.2016.015

[28] J. Zhang, F. Wang, K. Wang, W. Lin, X. Xu, and C. Chen. Data-driven intelligent transportation systems: A survey. *IEEE Trans. Intell. Transp. Syst.*, 12(4):1624–1639, 2011. doi: 10.1109/TITS.2011.2158001

[29] F. Zhao and X. Zeng. Optimization of transit route network, vehicle headways and timetables for large-scale transit networks. *Eur. J. Oper. Res.*, 186(2):841–855, 2008. doi: 10.1016/j.ejor.2007.02.005