Journal of Computer Languages 53 (2019) 1–8

Contents lists available at ScienceDirect

Journal of Computer Languages

journal homepage: www.editorialmanager.com/cola/default.aspx

View Points

Knotted-line: A Visual explorer for uncertainty in transportation system

Weixin Zhao[a], Hongyu Jiang[a], Kai Tang[a], Wenqi Pei[a], Yadong Wu[a,b,*], Abdul Qayoom[a,c]

[a] School of Science and Technology, Southwest University of Science and Technology, China
[b] Sichuan Civil-Military Integration Institute, Southwest University of Science and Technology, China
[c] Department of Computer Science, Lasbela University of Agriculture, Water and Marines Sciences, Pakistan

ARTICLE INFO

Keywords:
Uncertainty visualization
Public transportation system
User-centered design

ABSTRACT

Public transport system(PTS) is a complex system, and there are various uncertainties caused by traffic congestion, varying departure interval of buses, etc. In this paper, taking bus as an example, a novel visualization model knotted-line for transportation system uncertainties is proposed. The knotted-line visualization model consists of three parts, the violin plot, multi-layer ring and risk indicator. The violin plot shows the uncertainty of passengers' waiting time. The multi-layer ring describes the uncertainty of bus arrival time. The risk indicator expresses the possibility of passenger spending time on the trip. A user study is conducted to evaluate the proposed visualization method. The results show that the proposed knotted-line visualization model is helpful for passengers to understand the uncertainty in transportation system and help users to make decisions.

1. Introduction

Uncertainty refers to situations that contain incomplete or unknown information. Uncertainty factors can be used to describe uncertainty. Public transportation system(PTS) is influenced by uncertainty factors such as dynamic traffic operation, passenger requirements and so on, which makes passenger waiting time a random variable. Generally, passenger travel time is influenced by the passenger arriving time, bus schedules, traffic condition, weather factors, etc. Uncertainty is cumulative. The more passengers pass through the stations, the greater the uncertainty of travel.

The bus is one of the traffic tools in the public transportation system. By instantiating uncertainties, users can deepen their understanding of uncertainties [1]. Data visualization is an effective method to represent the distribution and structure of data sets and reveal the potential patterns of data. Visualization techniques provide a possibility to explain uncertainty. If there is no data visualization technology to help explain uncertainty, there will be more uncertainty in the reasoning process. Taking the bus system as an example, the knotted-line visualization model is designed to describe the uncertainty in the transportation system. The purpose of knotted-line is to convey the uncertainty in the transportation system to the users. The historical driving data of buses can be used to predict bus uncertainty information such as waiting bus time and travel time of passengers. Moreover, bus uncertainty information can be used to assist users in decision-making, helping users understand the unexpected and uncertain events in the bus system and improving service reliability. Taking the bus system as an example, this work discusses the uncertainty and visualization methods in the transportation system. Our work has the following contributions:

- **User Study:** To apply these results to a user-centered uncertainty visualization, users' goals need to be understood to gather the requirements in Mianyang, China.
- **Modelling:** To describe the uncertainty in bus system, this work used the history running data of bus to model the arrival time of buses and travelers waiting time.
- **Visualization:** With the requirements of travellers, the visualization model knotted-line is designed to present the overall information. Besides, some particular cases have been considered, such as bus transfer. The users can obtain the risk of schedules and uncertainty via the knotted-line model which can assistant users decision-making. Moreover, the safest riding scheme can be selected interactively through the knotted-line model.

2. Related work

In this section, we will review the related work, including visualization on decision making, uncertainty visualization, transportation visualization, and uncertainty visualization of transportation.

* Corresponding author to: Southwest University of Science and Technology, No.59, Qinglong Road, Fucheng District, Mianyang 621010, China.
E-mail addresses: weixin.zhao@outlook.com (W. Zhao), hongyuj@outlook.com (H. Jiang), wyd028@163.com (Y. Wu).

https://doi.org/10.1016/j.cola.2019.01.001
Received 3 May 2018; Received in revised form 31 December 2018; Accepted 2 January 2019
Available online 13 March 2019

2.1. Visualization on decision making

Poor data can affect the correctness of decision-making. Data visualization technology can be used to analyze the implicit data information in the visualization model, understood the potential uncertainty relationship of data and eliminate the impact of poor quality data on decision-making. Discussing data uncertainty and using uncertain information as supplementary information can enhance users' trust and decision-making in their daily environment. People can make better decisions when they are given point estimates with probabilistic information. Probabilistic information provides a more transparent form of information for decision-making, which leads to a higher trust of users. Joslyn & LeClerc [2] found that incorporating uncertain information in weather forecasting can increased user trust. There are some examples of how visualization affect the decision making. The different font sizes of words in the word clouds are proportional to the potential meaning. L.Rogers et al. [3] used statistical sampling to convey benefits and harms of treatment to doctors and patients. Then they can judge the effect of treatment. Garcia et al.[4] used visualization methods to help people understand health risks in order to make informed decision-making.

2.2. Uncertainty visualization

Uncertainty visualization is a tool for describing uncertainty. Uncertainty visualization methods include the iconography, visual variable coding, geometry representation, and animation representation [5]. Because there is no universal visualization method, the various visualization methods have unique applicable scenarios. Similarly, different uncertainty visualization models can express different uncertainty visualization goals. In order to make users perceived uncertainty, Pang et al. [6] encoded uncertainty information into attributions such as color, size, position and so on. In data visualization, different shapes and colors are used to express the degree of uncertainty, such as red for high uncertainty and green for low uncertainty. Grtler et al. [7] encoded a kind of new circular tree diagram to hierarchically structured data and its uncertainty. This visual model delivers uncertainty information through multiple visual variables. Uncertainty visualization focuses on presenting data together with auxiliary uncertainty information. The ultimate goal of uncertainty visualization is to provide users with visualization to help data analysis and decision-making.

2.3. Transportation visualization

With the development of society, the rapid growth of traffic poses great challenges to data analysis. Transportation visualization makes it possible for users to analyse traffic data. In transportation visualization, there is a linear relationship between bus routes and stations, which is suitable to be represented by a linear visualization model [8]. Ferreira et al. [9] provided a wide range of flexible method to querying time and space, filtering data segmentation and exploring OD data. This method provided a solution for exploring large-scale spatiotemporal OD data through visualization. Taking urban traffic as an example, Shimizu et al.[10] proposed a solution for the time-distance mapping. This solution can be applied to urban internal traffic problems, such as visualization of differences in transport service levels between peak and off-peak hours. Pu et al. [11] designed interactive visual analysis system based on taxi track which used to monitor and analysis of complex traffic condition in cities.

Uncertainty prediction is considered as the main purpose of uncertainty analysis [12,13]. How to present the prediction result is a big issue for researchers. Data-driven ITS improve traffic system performance by accessing and using traffic data [14]. Ferris et al.[15] described a set of transit tools named OneBusAway which focused on providing real-time arrival information for Seattle-area bus passengers. Wal et al. [16] designed three visualization models to help people explore and visualize passenger mobility in the public transportation system. Nguyen et al. [17] used the three-dimensional Cartesian coordinate system to represent the time map of moving objects. They put a temporal map to show the relationship between bus routes and time in space-time cube. The spatiotemporal cubes describe the traffic data that change location over time. Passengers can mark the bus trip for their more suitable travel.

2.4. Uncertainty visualization of transportation

Discussing the uncertainty visualization of transportation can assist us understand the uncertainties of transportation in a more intuitive way, which is conducive to improving transportation efficiency and decision-making reliability. Jung et al. [18] found that compared with single point estimation, displaying the range of residual power of electric vehicles as gradient map and fuzzy range can reduce the range of anxiety in driving tasks and increase the driver's trust in vehicles. They explored the effects of drivers' attitudes and driving behaviors on the display accuracy of different instrument estimation ranges and charging states under different resource availability conditions. Kay et al. [19] proposed a novel mobile interface design and visualization of uncertainty for transportation predictions on mobile phones based on discrete outcomes. Wunderlich et al. [20] proposed a visual design method to show the scheduled train connections and the expected train delays. Uncertainty visualization can describe situations that are unclear or involve unknown information. Uncertainty is divided into two categories, including cognitive uncertainty and arbitrary uncertainty [21]. The uncertainty of buses belongs to cognitive uncertainty. The reason is that the information about travellers and buses are not symmetrical and it is difficult to predict the operation of buses. Therefore, uncertainty visualization of transportation should be based on traveller's cognition to describe bus uncertainty, helping passengers understand the uncertainty of the public transportation system and make travel plans.

3. Task analysis

To clarify the task requirements and achieve the effective visual design, we summarized the most concerned problems about bus journey and discussed the uncertainty and visual design of the bus.

3.1. Passenger goals

To follow user-centered visual design guidelines, we need to capture and understand user goals. To this end, we conducted a survey to identify 1) How do the users arrange their trip schedule and 2) What factors that users need to consider when they plan a trip. We surveyed 86 residents in Mianyang, China, through social media and e-mail. The top 4 highest-rated questions users currently asks are:

- **Wait time**: If I leave now, how long do I have to wait at the bus stop?
- **In-vehicle journey time**: If I get on the bus now, how long can I arrive at the destination?
- **Status probability**: The passengers want to know the probability of arrival after a certain time point and within a certain time period.
- **Schedule risk**: Will I get to a meeting/event on time despite bus delays? It relates to a commonly-described worst experience of buses coming later than expected.

3.2. Task-oriented analysis

According to passenger demand, five demand-oriented analysis objectives are summarized and visualization methods to achieve analytical objectives are discussed.

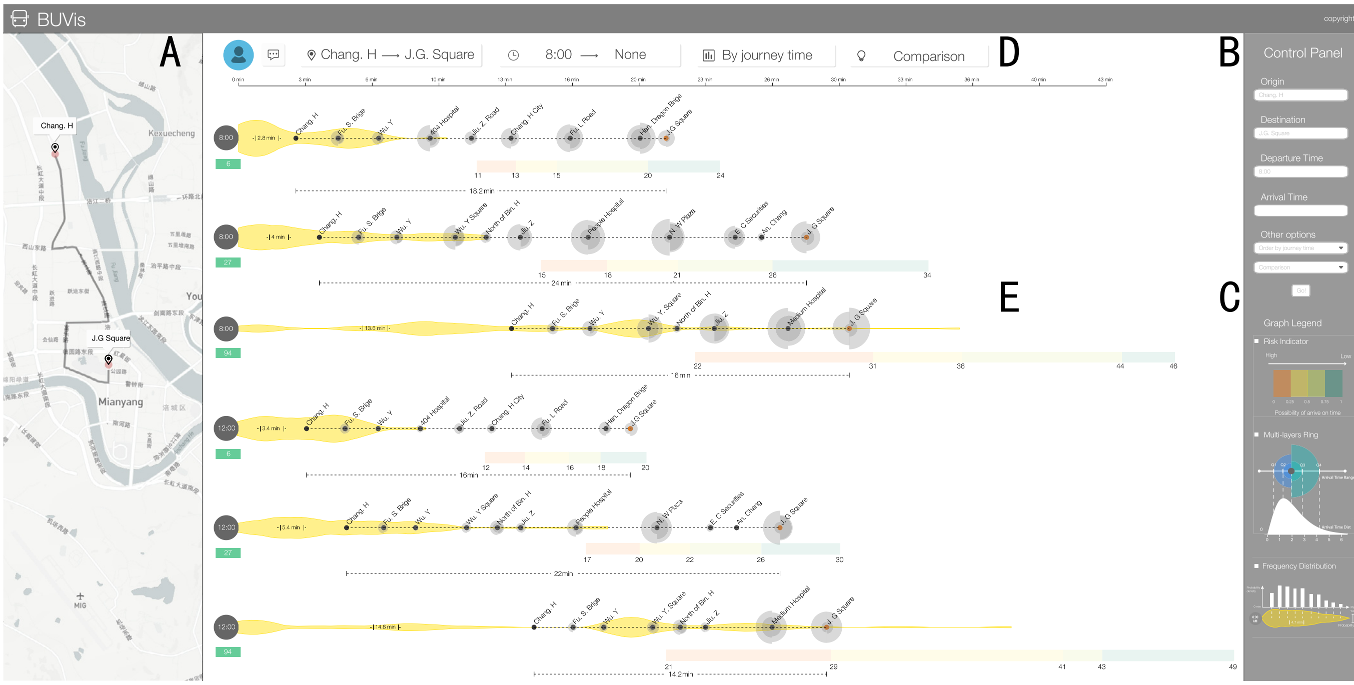

Fig. 1. An overview of bus uncertainty visual analysis system. (a) The map includes basic information such as recommended routes and departure and arrival stations. (b) Users can set up basic visual configuration information such as stations, time, resulting arrangement, etc. through the control panel. (c) This panel is a visual component legend to help users understand uncertainty. (d) Information bar shows users' brief configuration. (e) Main view is the visualization model of the knotted-line. According to the user configuration, the whole bus journey is shown with elements of uncertainty, and the optimal choice is obtained by comparison.

- **Schedule Evaluation** For each schedule, the cumulative probability for the bus wating time is calculated. The higher the cumulative probability, the safer the travel plan and the lower the risk of travel. In our visualization, as shown in Fig. 1, we can use color to map the risk of passenger itinerary planning. Rectangular length in the knotted-line model mapping to complete time range of the itinerary. Passengers can then analyze the success rate of the itinerary plan.
- **Transfer scenarios** Generally, passengers need to take one or more buses to reach their destination, which involves the transfer behavior. Visual models need to support multiple user behaviors. In our visualization, component-based visual models are used to satisfy multiple user behavior scenarios. For example, we will draw multiple timelines to represent multiple bus lines because of the transfer situation. As shown in Fig. 5, timelines are connected end to end and time-cost will be composite.
- **Comparison** Most of the travellers hope to use the minimal time to reach the destination. Travellers can arrive at their destination with N kinds of itinerary plans. The success rate of each travel plan depends on the waiting time of the passengers, the running time of the bus and passengers' flexible time. In each bus line, more stops does not mean less time. In short-distance travel, some important transport hub stations in areas with high traffic load may also spend much time. Therefore, our visualization models need to support users to analyze the time cost of different routes and determine sites with high uncertainty. The visualization system is shown in Fig. 1. Passengers start at 8 o 'clock and set 4 hours as the time granularity to check the bus from Chang. H to J.G Square. The results show that the recommended routes can be divided into No. 6 bus, No. 27 bus and No. 94 bus. The No. 6 bus is the best choice in the whole plans because of the shortest speed time of travel as shown in the graph. Although No. 94 bus is superior to No. 6 bus in the number of stations and travel time, the long waiting time and fluctuation of arrival time of No. 94 bus make the whole journey too uncertain and unsuitable to take.
- **Schedule Recommendation** Usually, the user needs to know the best departure time after determining the itinerary. As shown in Fig. 7, users can choose the expected time to arrive at the destination. The module can recommend departure time ranges with different risk levels for users. The users can select a time to departure according to the importance of schedules.
- **Road Condition Assessment** The running state of buses can reflect the traffic condition. The running data of buses are sampling with a fixed cycle, which indirectly included the transportation information of the city. Hence, the running stability of buses can reflect the traffic conditions in an area. For example, the unstable arrival time of a bus at a station may be due to the high uncertainty of traffic conditions near the station, such as traffic jam caused by the increase in traffic flow.

4. Data description

4.1. Data introduction

The experimental data were provided by Mianyang City Bus Company. Mainyang is a third step city in China, where residents mainly travel by bus [22]. The data is based on GPS bus track with a data size of 4.5 GB and a time span of one year from 2016-1-1 to 2016-12-31. The raw data contains bus running state such as bus arriving time, bus departure time, longitude and latitude, bus number, road sections, id, etc. The stability of each bus data recording system is different, which may record noise data. The original data need to be pre-processed in advance for subsequent calculations.

4.2. Data processing

The raw data is composed of continuous GPS spatio-temporal information. But it does not include site information such as site name, site longitude and latitude. Raw data need to be matched with site information according to GPS position. We hold the opinion that the shortest distance between the original data and the location to be matched is the target matching point. Latitude and longitude are in the coordinates of the spherical system rather than cartesian coordinates.

The distance between two coordinates cannot directly be calculated by latitude and longitude. We use the Haversine formula to work out the camber distance between two coordinates and match corresponding data:

$$\alpha = \varphi_1 - \varphi_2$$
$$\beta = \lambda_1 - \lambda_2$$
$$d = 2r\cdot arcsin\left(\sqrt{sin^2\left(\frac{\alpha}{2}\right) + \cos(\varphi_1)\cos(\varphi_2)sin^2\left(\frac{\beta}{2}\right)}\right) \quad (1)$$

The d is the camber distance between two coordinate points. The r is the earth's average radius. Besides, the φ_1 and the φ_2 correspond to the corresponding latitude of coordinate points A and B. Also, the λ_1 is the longitude of point A and the λ_2 is the longitude of point B. The threshold of calculation results is set to improve site matching performance in the process of calculation. We hold the opinion that it is the error data recorded by the bus system which should be deleted when the calculated result is greater than the threshold value. After the above operations, site information and raw data will be matched. The bus arrival time and departure time at each station are extracted. This makes it possible to calculate the time interval of any adjacent pair of stations. We can get any combination of stations to support the exploration of complex situations, such as passengers transfer bus. It provides data support for the exploration of complex situations. Besides, because the bus will stop for a long time after it arrives at the terminal, GPS will keep recording the same location information. This part of invalid data should be removed to reduce data volume and save storage space.

4.3. Bus arrival time model

There is a passenger arriving at the bus station at time t. He wants to take the nearest bus $\phi(t)$. The point estimation and probability estimation of bus arrival time can be calculated using the dataset $\{W_t^{\phi(8:00)}, t \in T\}$ of bus arrival time on the same working day.

The queuing theory is a subject that studies the random law of queuing phenomena in service system. Following the queuing theory, if the time point of bus arrival is an update process, namely, the interval of depart is independent [23]. Then the expectation of passengers' waiting time can be calculated by the following equation:

$$h_f = \frac{\lambda^{-1}}{2}(1 + C^2) \quad (2)$$

Among them, λ^{-1} is the mean of the bus depart interval. C is the variation coefficient which can be calculated with (*std/mean*). The variation coefficient C is usually used to measure the randomness degree of a non-negative random variable. If the depart interval is fixed, the variation coefficient C is zero which means there is no randomness. Poisson process is a kind of stochastic process, which is defined by the occurrence time of events. It is the most basic independent incremental process for accumulating the occurrence times of random events. In contrast, if the statistical growth process of the number of bus departure intervals is a Poisson process, the variation coefficient C is equal to one which means it is random. However, in the real world, bus departure intervals are fixed or random.

Hence the variation coefficient C is between 0 and 1. It means the interval time of bus occurrence fit the Erlang distribution and the bus arrival time is relatively smooth. Of course, in some extreme cases, the variation coefficient may be beyond 1. For example, there are 100 buses arriving at their destination at the same time, resulting in long-term bus outage, which is usually called sudden arrival. Although the waiting time is unlimited in theory, the passenger will choose another way to reach their destination after exceeding the expected waiting time. The gamma distribution is a property distribution composed of two parameter families of continuous probability distribution in probability theory and statistics. It is often used to model waiting time. From the above discussion, it can be seen that the departure intervals of buses obey the Erlang distribution in real life. The Erlang distribution is a special case of the gamma distribution. If the shape parameter k of the gamma distribution is a positive integer, the gamma distribution represents the Erlang distribution. Because passenger waiting time and bus arrival time are based on bus departure interval, the gamma distribution can be used to fit the passenger waiting time and bus arrival time and draw the probability density graph. Gamma distribution can be parameterized with reference to shape parameter α and inverse scale parameter β. The shape parameter α is equal to the quadratic of the inverse variation coefficient of departing interval $1/C^2$, while the inverse scale parameter β can be calculated with $mean/std^2$. The *mean* is the average waiting time of passengers. The *std* is the standard deviation of passenger waiting time. In the real world, the bus departing interval approximates the mean of arriving time. Hence, the gamma distribution can be used to fit the bus waiting time. The gamma distributed with the bus waiting time can be denoted by:

$$\Gamma(\alpha, \beta) \equiv Gamma(mean^2/std^2, mean/std^2) \quad (3)$$

The corresponding probability density function (PDF) in the shape-rate parametrization given by

$$f(x; \alpha, \beta) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)} \quad for\ x > 0\ and\ \alpha, \beta > 0 \quad (4)$$

The cumulative distribution function (CDF) of gamma distribution $F(x; \alpha, \beta)$ is calculated by [24]:

$$F(x; \alpha, \beta) = \int_0^x f(u; \alpha, \beta)\, du = \frac{\gamma(\alpha, \beta x)}{\Gamma(\alpha)} \quad (5)$$

where $\gamma(\alpha, \beta x)$ is the lower incomplete gamma function. The cumulative distribution function (CDF) is used to calculate the summarize the probability of all possible situations in random events. It is used to illustrate and communicate the point estimation and probability of waiting bus time.

5. Visual design

In order to explore the appropriate visualization expression method of uncertainty and describe bus uncertainty by visualization method, the basic design of visualization model and its principle are discussed in this section. We describe the uncertainty visualization of the bus from part to whole.

Firstly, calculating the cumulative distribution of passenger waiting time to provide point estimation and probability estimation. Secondly, the multi-layer ring is used to divide the cumulative probability distribution of bus arrival time into several time ranges. The multi-layer ring can better reflect the probability of bus arrival and the uncertainties of a station in the corresponding time range. Then, according to the theory of four primary colors in color psychology and the probability distribution of passenger's travel completion time, the risk indicator of passenger's travel complete time is constructed. Different colors correspond to the possibility of arriving at different time ranges, with the aim of presenting an overview of the whole trip.

5.1. Visual design overview

Geometric elements are an important part of visual design works. User's cognition will be affected by the characteristics of geometric elements and different ways of expression, such as different arrangements and combinations. To obtain an appropriate visualization model, metaphors of geometric figures and applicable scenarios need to be considered. The rectangle shape is reliable which give stability and suggest the order. The straight-line shape and right angle shape confer a risk range of reliability and safety. Besides, circles represent completion, wholeness, and harmony. Circles convey the idea of boundaries. Violin shape gives a sense of continuity to the flow of people.

According to Gestalt psychology, the eye and brain do not

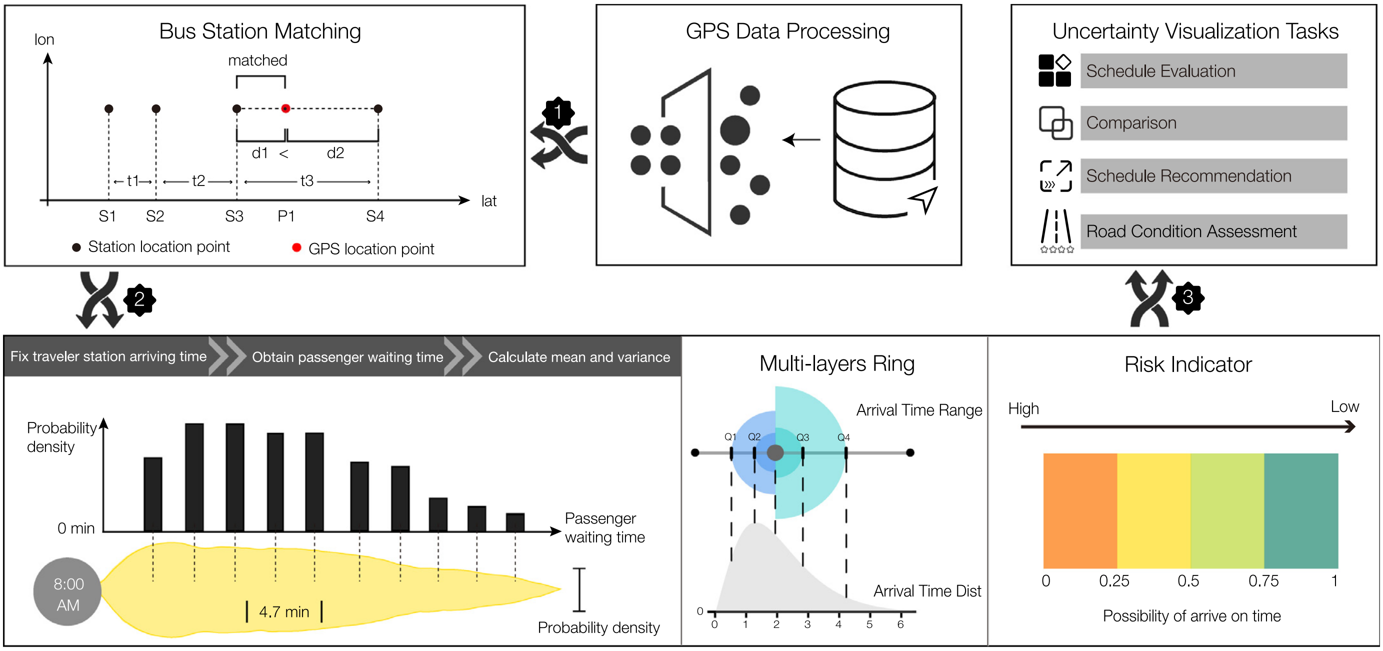

Fig. 2. The pipeline of uncertainty visualization framework about bus data. Data processing is the first step which include operations such as removing noise data and so on. Because there is no site information in our data, the algorithm of station matching via curve distance will be used to matching the station with maximum probability. Then, the history waiting and arrival time will be used to build the various uncertainty visualization components. Finally, based on different requirements, the uncertainty visualization model knotted-line is used for task analysis.

distinguish components of a single component of an image in the first place but combine parts to make a more comprehensible unity when people are watching. According to the theory of four primary colors in color psychology, mental reactions and behaviors to different colors are different. Psychological activity can be influenced by color. In our visual design, the timeline is regarded as a basic visual element. Different combinations of geometric figures are used to express specific meanings according to different scenarios. As shown in Fig. 2, according to the characteristics of rectangle length range, risk indicator indicate the possibility of passenger travel time. Besides, using the characteristics of the arc radius boundary to express the possibility of bus arrival in the corresponding period. Moreover, violin shape is used to fit the point and probabilistic estimation of passenger waiting time.

5.2. Knotted-line

According to our investigation, the traveller hopes to obtain the information about bus time. Also, they want the visualization tools to be simple and intuitive. Center on this goal, the basic visual elements are chosen to make up the knotted-line, making it practical to improve the cognitive efficiency of user. We combine the violin plot, multi-layer rings and risk indicator in one line into a visualization model named **knotted-line** which is used to present the whole uncertainty information in a line. As shown in Fig. 3, yellow background graph is the violin plot, which is the cumulative distribution graph fitted by passenger's waiting time data. The violin plot can provide the point estimation and probability estimation of the waiting time. The time marked in its area is the average waiting time. The black dots on the knotted-line are different bus stops whose location is determined by the average time of arrival at the bus stop. The black dot is surrounded by a multi-layer ring, which visualizes the cumulative distribution of bus arrival at each station. It supports the point estimation and probability to present uncertain information. The second dotted line marks the average running time of the bus. Users can avoid the high uncertainty of the station, choose the appropriate ride plan, and evaluate the travel time cost to achieve the purpose of using uncertain information to assist decision-making.

The length of the knotted-line is calculated by the average journey time which is summed by arriving time s_t of each station s, $s \in (Origin, Destination]$. To fit the width of the display window W, the result will times a ratio factor f_r. Hence the f_r is given by:

$$\sum_{s \in (O,D]} s_t \times f_r / W = 1 \tag{6}$$

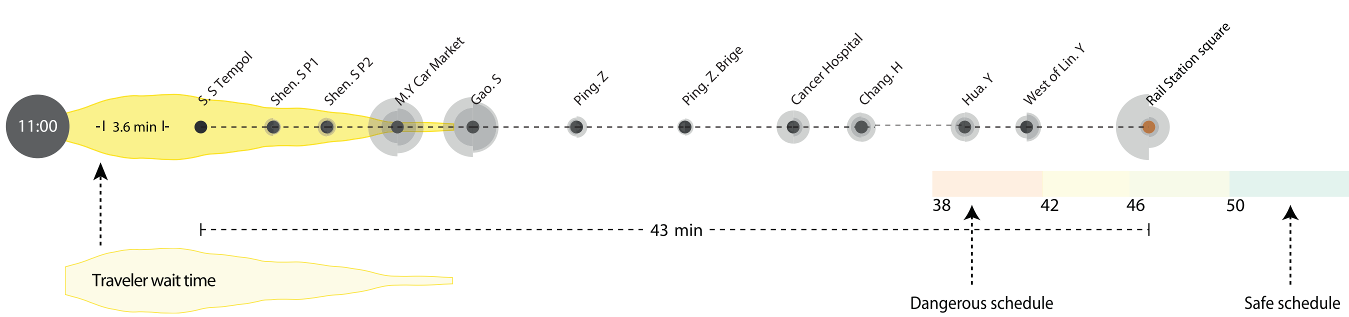

Fig. 3. The design scheme of the knotted-line. Passengers can estimate the waiting time and bus arrival time by point and probability through violin plot and multi-layer ring. Besides, risk indicator is used to perceive travel safety. It can describe the uncertainty and time-cost of the whole bus journey.

In addition, in the union model knotted-line, different colors are used to indicate the success probability of scheduling.(as shown in Fig. 3)

For example, there is a girl Ann. She is at the Q.Y station at 11:00 AM. She wants to take the train which will depart at 12:10 PM. She should arrive at the railway station at 12 PM. According to the above presupposition, we get the knotted-line model as shown in Fig. 3. As can be seen from Fig. 3, she needs to take No. 27 bus and No. 38 bus and transfer at S.S Temple. She had to wait two buses at the departure place and transfer point, with an average waiting time of 7.3 minutes. According to multi-layer ring, stations with high uncertainty have S.S Temple, M.Y Car Market, Gao. S and Rail Station square. The risk indicator is used to predict travel risk. As shown in Fig. 3, the probability of completing the journey in 55 minutes is more than 95%, which shows that her journey is safe.

5.3. Violin plot

Although real-time arrival time estimation applications are often used to quickly determine when to arrive at bus stops [19], point estimation without uncertainty often conveys error accuracy. Probability estimation will help users understand the possibility of bus arrival within the target period, which will help people assess schedule risks and schedule opportunities. Probability estimation is related to scenario risk tolerance, which helps people to estimate risk conservatively according to different scenarios. The combination of point estimation and probability estimation will help people better answer when to leave, waiting time and schedule risk. Get people ready in advance to cope with the commonly-reported worst experiences, such as the unexpected early or late arrival of buses.

Hence, we use the violin plot (Fig. 4) to illustrate the point estimation and probability estimation of traveller's waiting time. The information in the middle of the violin plot is the point estimation of the average waiting time of passengers. The probability estimation is provided by the proportion of violin area in different periods. They provide basic awareness of bus waiting time uncertainty. Yellow is used as the background color of the violin plot that can give hints to people.

To build it, we use the history data W which is comprised of the bus waiting time W_t each day. The data of passenger waiting time are used to calculate the mean and variance. The cumulative distribution (Formula 3 and 5) is calculated by the mean and variance of waiting bus time. As shown in Fig. 4, in the cumulative distribution function plot, the horizontal axis is the passenger waiting time, and the vertical axis is the probability density. We can estimate the probability of bus arrival in the corresponding period by the percentage of the graphics area. In order to better understand the violin chart, we use the histogram to help understand the violin chart. We create a histogram with fixed-number bar H to illustrate the probability density of waiting time data for each group of buses.

Each bar denotes the probability density H_p in the corresponding range $\{r, r \in [min, max]\}$. Given the height of view *Height*, the height of base points in the violin plot is:

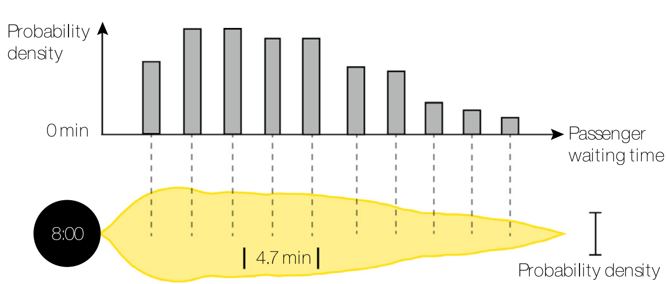

Fig. 4. The frequency histogram of wait-time was translated to a violin plot which presenting the point estimation and probability, the label in the violin shows the average value of wait-time.

$$H_i = \frac{F_i - F_{min}}{F_{max} - F_{min}} \times Height \tag{7}$$

5.4. Multi-layer ring

People can better understand probability information by dividing probability into multiple probability intervals. The multi-layer ring model is used to express the arrival time and probability intervals of buses. Because the bus departing interval approximates the mean of arriving time in the real world, the cumulative distribution (Formula 3 and 5) is calculated on the basis of the mean and variance of bus arrival time. Fig. 6 shows the cumulative distribution of bus arrival time at the station. In the cumulative distribution function plot, the horizontal axis is the arrival time of the bus, and the vertical axis is the probability density. The total cumulative distribution plot area is 1. According to the positive direction of the transverse axis, the values Q1, Q2, Q3 and Q4 are determined when the image area are 0.1, 0.3, 0.7 and 0.9.

The values Q1, Q2, Q3 and Q4 above the multi-layer ring indicate that the probability of bus arriving before the corresponding time are $P(t) = 0.1, 0.3, 0.7$ and 0.9. Any two points on the multi-layer ring can be selected to describe the time range and probability of arrival. For example, the bus arrival probability of corresponding time range between Q1 and Q4 is 0.8.

5.5. Risk indicator

Human psychology is influenced by color. Color is regarded as a kind of stimulus which can produce specific emotions and excite unintentional judgment and expectation. In color psychology, colors are divided into four types: red, yellow, green and blue, which called the four primary colors of psychology [25]. The response of the brain to red is vigilance, and red stimulate the visual nerve and impulses the mind. Besides, the reaction of the brain to blue is relaxation and blue makes our heart calm. Correspondingly, we use four primary colors to express the different risk levels of the bus arrival time in Fig. 2 named risk indicator. The probability distribution (Formula 3 and 4) is calculated on the basis of the mean and variance of bus arrival time. The length of each block is positively correlated with probability. From left to right of the risk indicator, the greater the probability of completing the journey, the less dangerous it is. Different colors correspond to different time range of bus arrival probability. From red to blue, the probability of successful bus arrival increases and the risk of travel arrangement decreases. Users can able to get an overview of how long the trip took from risk indicator, giving users an overview of the whole trip.

6. Experiment and evaluation

In this section, we will introduce the experimental results to discuss the effectiveness of bus uncertainty in assistant time estimation. We first introduce the participants and the experimental process. Then different estimation methods are used to compare stability. Finally, potential factors affecting estimation accuracy are analyzed.

6.1. Participants

Before the experiment begins, we need to determine the preset experimental route for the experiment. Fixed originating station, according to the number of passing stations, three different stations were chosen as terminal stations, which constitute a total of three routes. By the different target route of passengers, we choose the passengers who correspond to the scheduled route as participants in this experiment. Three experimental routes divided participants into three groups, a total of 10 participants. The participants were city residents, ranging in age from 20 to 28, with different occupational and educational backgrounds. Due to the limited number of samples, the influence of

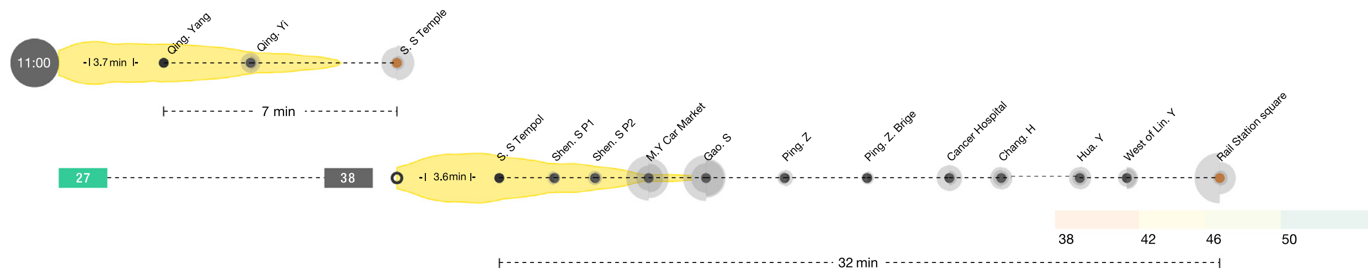

Fig. 5. The transfer case of the knotted-line, a special case for reach destination should reach to S.S Temple via line 27 then transfer to line 38. And the start time of the second line is the end of the first line.

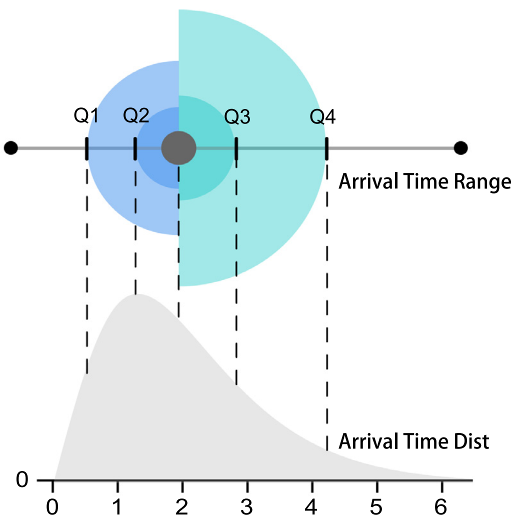

Fig. 6. The multi-layer ring, it was used to present the point estimation and probability of bus arrival time.

professions and educational backgrounds are not considered.

6.2. Procedure

In addition to the length of the journey, the passenger travel time is also affected by the uncertainty factors accumulated with the number of stations. During the experimental process, firstly, participants estimate the total travel time consumption based on experience, which is the sum of the waiting bus time and taking bus time. Secondly, participants learn the uncertainty visualization model knotted-line to estimate the completion time of the whole trip. Then, we follow participants record the actual time consumption. As a result, we got three kinds of data, user experience estimation time, knotted-line estimation time and actual travel time.

6.3. Result and analysis

We compared the experimental results of different groups and discussed the influence of uncertain information on assistant decision-making from different estimation methods and the different number of sites. The analysis shows that it is effective to use the knotted-line model to help users understand bus uncertainty. Using knotted-line models can help users make decisions. It can let users know the overview of the journey, help passengers make travel plans, assess the risk of the journey, reduce passengers' waiting time for buses, and play an active role in improving travel efficiency.

6.3.1. Deviation comparison between different estimation methods

We compared two estimates methods: users experience estimation time and knotted-line estimation time. The actual measurement time is the recorded travel completion time. The deviation between the actual measurement time and the two estimation time is calculated to illustrate the stability difference between the two estimation methods. As we can see in Fig. 8, among the three groups of A, B and C users, The fluctuation of time deviation estimated by the knotted-line model is smaller than that estimated by experience. It shows that the time range of experience estimation is larger than the time range of knotted-line estimation. Therefore, the time can be estimated more accurately by using the knotted-line model, which can improve the travel efficiency of users. Similarly, it can be seen that the use of bus uncertainty information to assist decision-making can improve the accuracy of decision-making.

6.3.2. Influence of different station number

We analyzed the impact of the number of stations on the accuracy of passenger estimation time. Generally, the more stations and the longer distances, the more uncertainties they accumulate, and the greater the impact of uncertainties on bus driving. Passengers find it difficult to grasp the impact of changes in the number of stations on time estimation. So the results of passenger experience time estimation are unstable. The experimental results showed in Fig. 9. Among the three groups of A, B and C users, the fluctuation of user's experience estimation time increases with the increase of the station number.

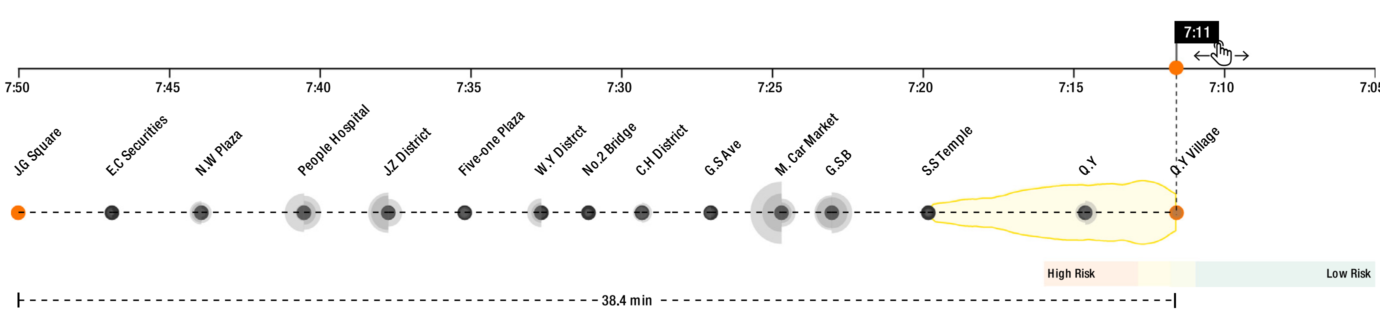

Fig. 7. Schedule recommendation, the users can select an exception arrival time for his schedule, then our system will recommend a series departure-time ranges to deal with the different importance of the schedule.

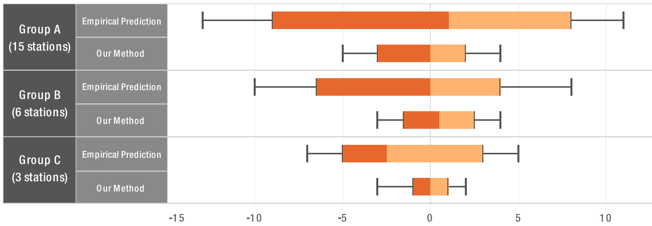

Fig. 8. The box-plot of three group experiments with two estimation method. The rectangles in each survey group with deep color indicate the arrival time later than expected, while the light color indicates the arrival time early than expected.

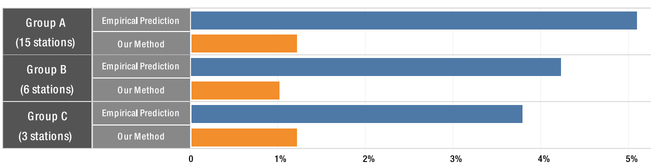

Fig. 9. The deviation between the two estimation methods and measured values. It is observed that the estimated value deviation of our method is less than the empirically estimated value deviation and is not affected by the number of stations.

Participants use the knotted-line to estimate the time is more stable. This suggests that the knotted-line model has a positive impact on passenger time estimates. Using uncertain information to assist decision-making can improve the accuracy of decision-making. Also, it is difficult for most users to make an accurate prediction based on accumulated travel experience before using the knotted-line model. After the completion of the experiment, most users think that using the knotted-line model are helpful to their travel prediction.

7. Conclusion

In this work, taking the bus as an example, we discussed the uncertainty and visualization method in the transportation system. According to Gestalt psychology, color psychology and geometric figure metaphor, the knotted-line visualization method, including the violin plot, multi-layer ring and risk indicator, is given to describe the uncertainty of waiting time and travel time. Users can use the bus uncertainty information in knotted-line model to assist decision-making. Finally, a user study is conducted to evaluate the validity of the knotted-line model. In the future work, we will improve the effectiveness of our method, and enhance the prediction algorithm and design visualization model with stronger interpret-ability.

Acknowledgements

The research of authors is partially supported by National Natural Science Foundation of China (No. 61872304, No. 61802320, No. 61872066, No. 61502083), National Key R&D Program of China (2016QY04W0801) and the National Defense Program on Basic Research Project (No. JCKY2017404C004).

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cola.2019.01.001.

References

[1] M. Brehmer, T. Munzner, A multi-level typology of abstract visualization tasks, IEEE Trans. Vis. Comput. Graph. 19 (12) (2013) 2376–2385.
[2] S. Joslyn, J. LeClerc, Decisions with uncertainty: the glass half full, Curr. Dir. Psychol. Sci. 22 (4) (2013) 308–315.
[3] L. Rogers, et al., The cochrane collaboration, Cochrane Database of Systematic Reviews, UK, 2009.
[4] R. Garcia-Retamero, E.T. Cokely, Communicating health risks with visual aids, Curr. Dir. Psychol. Sci. 22 (5) (2013) 392–399.
[5] S. Deitrick, R. Edsall, The influence of uncertainty visualization on decision making: an empirical evaluation, Progress in Spatial Data Handling, Springer, 2006, pp. 719–738.
[6] A. Pang, et al., Visualizing uncertainty in geo-spatial data, Proceedings of the Workshop on the Intersections between Geospatial Information and Information Technology, National Research Council Arlington, VA, 2001, pp. 1–14.
[7] J. Görtler, C. Schulz, D. Weiskopf, O. Deussen, Bubble treemaps for uncertainty visualization, IEEE Trans. Vis. Comput. Graph. 24 (1) (2018) 719–728.
[8] A. Pang, C. Wittenbrink, S. Lodha, Approaches to uncertainty visualization, Vis. Comput. 13 (8) (1997) 370–390.
[9] N. Ferreira, J. Poco, H.T. Vo, J. Freire, 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) (2013) 2149–2158.
[10] E. Shimizu, R. Inoue, Time-distance mapping: visualization of transportation level of service, Proc. of Symposium on Environmental Issues Related to Infrastructure Development, (2003), pp. 221–230.
[11] J. Pu, S. Liu, Y. Ding, H. Qu, L. Ni, T-watcher: A new visual analytic system for effective traffic surveillance, Mobile Data Management (MDM), 2013 IEEE 14th International Conference on, 1 IEEE, 2013, pp. 127–136.
[12] R. Moffat, Contributions to the theory of single-sample uncertainty analysis, J. Fluids. Eng. 104 (2) (1982) 250–258.
[13] M.W. Binford, Calculation and uncertainty analysis of 210 pb dates for pirla project lake sediment cores, J. Paleolimnol. 3 (3) (1990) 253–267.
[14] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, C. Chen, et al., Data-driven intelligent transportation systems: a survey, IEEE Trans. Intell. Transp. Syst. 12 (4) (2011) 1624–1639.
[15] B. Ferris, K. Watkins, A. Borning, Onebusaway: results from providing real-time arrival information for public transit, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, 2010, pp. 1807–1816.
[16] W. Zeng, C.-W. Fu, S.M. Arisona, A. Erath, H. Qu, Visualizing mobility of public transportation system, IEEE Trans. Vis. Comput. Graph. 20 (12) (2014) 1833–1842.
[17] H.T. Nguyen, C.K.T. Duong, T.T. Bui, P.V. Tran, Visualization of spatio-temporal data of bus trips, Control, Automation and Information Sciences (ICCAIS), 2012 International Conference on, IEEE, 2012, pp. 392–397.
[18] M.F. Jung, D. Sirkin, T.M. Gür, M. Steinert, Displayed uncertainty improves driving experience and behavior: the case of range anxiety in an electric car, Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, ACM, 2015, pp. 2201–2210.
[19] M. Kay, T. Kola, J.R. Hullman, S.A. Munson, When (ish) is my bus?: User-centered visualizations of uncertainty in everyday, mobile predictive systems, Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, ACM, 2016, pp. 5092–5103.
[20] M. Wunderlich, K. Ballweg, G. Fuchs, T. von Landesberger, Visualization of delay uncertainty and its impact on train trip planning: a design study, Computer Graphics Forum, 36 Wiley Online Library, 2017, pp. 317–328.
[21] K. Potter, P. Rosen, C.R. Johnson, From quantification to visualization: a taxonomy of uncertainty visualization approaches, Uncertainty Quantification in Scientific Computing, Springer, 2012, pp. 226–249.
[22] W. Pei, Y. Wu, S. Wang, L. Xiao, H. Jiang, A. Qayoom, Bvis: urban traffic visual analysis based on bus sparse trajectories, J. Visualizat. (2018).
[23] V. Sundarapandian, Probability, Statistics and Queuing Theory, PHI Learning Pvt. Ltd, 2009.
[24] A. Papoulis, Random Variables and Stochastic Processes, McGraw Hill (1976).
[25] A.J. Elliot, M.A. Maier, Color psychology: effects of perceiving color on psychological functioning in humans, Annu. Rev Psychol 65 (2014) 95–120.