Journal of Visual Languages and Computing xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Visual Languages and Computing

journal homepage: www.elsevier.com/locate/jvlc

The visual analysis of flow pattern for public bicycle system☆

Xiaoying Shi[a,b,*], Zhenhai Yu[a], Jing Chen[a], Haitao Xu[a], Fei Lin[a]

[a] School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, China
[b] Key Laboratory of Complex Systems Modeling and Simulation, Ministry of Education, Hangzhou, China

ARTICLE INFO

Keywords:
Visual analytics
Intelligent transportation system
Public bicycle system
Traffic data visualization
Human mobility

ABSTRACT

Public Bicycle System(PBS) is an increasingly popular mode of public transit, with the advantage of pollution-free and flexibility. In this paper, we present an interactive visual analytic system for exploring complex flows generated by PBS. Four inter-linked visualization views are designed to illustrate multiple perspectives of data, such as the spatial-temporal changes, the relationships and differences between flow OD pairs and the multi-dimensional factors(weather condition, calendar events) influencing on the rental numbers. A new presentation "Parallel coordinates with line and set" combined with flexible interaction schemes is proposed to support the exploration of multivariate association. We exemplify our approach with a real citywide PBS dataset. The results of case study demonstrate that our system is helpful for visually classifying stations with different flow patterns, speculating in-depth reasons, as well as investigating abnormal behaviors, helping decision makers to gain a better understanding of the large dataset.

1. Introduction

Public transportation systems play an important role in modern cities, which provide shared transportation services for the general public. Public Bicycle System [1–3] is an increasingly popular and green mode of public transit. For a city equipped with PBS, a lot of stations distribute around it. The users can pick-up or drop-off bikes through fixed docking slots by smart cards. Studying the flow pattern of this system is highly beneficial to users as well as the administrative department. However, since the users can choose rental route and time freely, the travel purpose contains high variability due to the different travel demand every day. How to adopt effective methods to support the exploration of inherent data features is challenging.

Visual analysis technology has been used to assist the analysis of PBS data. Some initial research works were devoted to inspect its characteristics by designing visual analytic systems [4–6]. Most previous works used simple charts to show data's spatio-temporal properties, and didn't analyze multiple factors impacting flow amounts, such as weather condition or calendar events. Corcoran et al. [7] studied multiple influences on the usage pattern of PBS. However, the analyzer didn't be endowed with power to explore flow patterns interactively through a user-friendly interface.

After discussing with the transportation administrator, we design a visual analytic system in this paper to perform three tasks:

Task 1 Categorizing stations and understanding the main function of one specific station, to provide advice for optimal bicycle dispatch;.

Task 2 Showing the flow correlations and differences among multiple stations, to explore the spatial direction of bicycle moving;.

Task 3 Visualizing the influence of multiple factors on bicycle rental numbers, to find the regular rules and abnormal events.

In order to achieve these tasks, four visualization views embedding spatial, temporal and multi-dimensional perspectives are designed to discover knowledge from dynamic and multivariable PBS data. In addition, when analyzing multiple factors affecting the rental numbers, one major challenge is the heterogeneity of attributes. For example, the temperature attribute is continuous, while the weather condition is discrete. The parallel coordinates [8,9] used for multi-dimensional visualization is not suitable for attributes having different qualities and would lead to overlapping. We propose a new presentation "Parallel coordinates with line and set(PCLS)", which adopts one-to-one item mapping for numerical attributes, and a frequency-based representation for categorical attributes, to support the exploration of multivariate association. We exemplify our approach with a real citywide PBS dataset. The system is able to help the transportation administrator investigate flow patterns, schedule the bicycle effectively, as well as find abnormal behaviors.

The main contributions include: 1) A visual analytic system is developed to study the features of PBS data from spatial, temporal and multi-dimensional perspectives, which offers a new way to inspect the dynamics of bicycle movements. 2) We propose a new representation

☆ Fully documented templates are available in the elsarticle package on CTAN.

* Corresponding author at: School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, China.
E-mail addresses: shixiaoying@hdu.edu.cn (X. Shi), 517191218@qq.com (Z. Yu), cj@hdu.edu.cn (J. Chen), xuhaitao@hdu.edu.cn (H. Xu), linfei@hdu.edu.cn (F. Lin).

http://dx.doi.org/10.1016/j.jvlc.2017.03.007
Received 15 September 2016; Received in revised form 22 January 2017; Accepted 23 March 2017

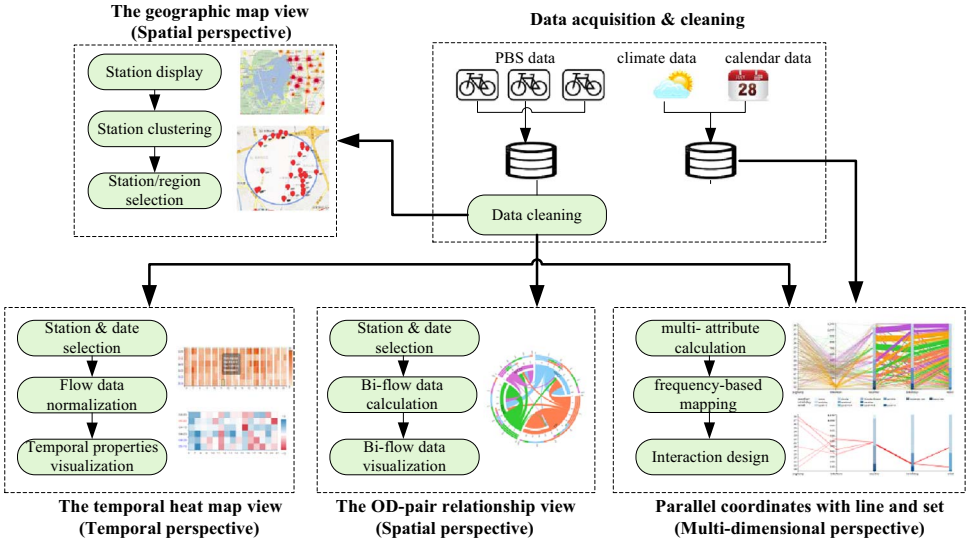

Fig. 1. System overview.

"Parallel coordinates with line and set", which supports the analyzer to observe the relationship of different heterogeneous factors interactively.

The paper is organized as follows. In Section 2, a brief overview of the related work is given. Section 3 describes the workflow of our system and the design of four visualization views in detail. Case studies are discussed in Section 4. Section 5 concludes the paper with suggestions for future work.

2. Related work

2.1. The visualization of traffic data

Traffic data contains human digital footprints on the space, which is hard to directly get knowledge from its original form. Visual analysis tools are designed to explore traffic data more conveniently [10].

Since the public bicycle system is a new kind of traffic mode, the research on the analysis of its underlying features is just in beginning. Mooney et al. [5] used simple line chart and histogram to analyze its lease purposes and flow pattern. Froehlich et al. [6,11] provided a spatio-temporal analysis of PBS data to discover the daily routines. Wood et al. [4] presented three visual representations to identify the travel behavior changing over space and time. The results of paper [12] indicated the cycle flows could be used in the planning of dedicated bicycle lanes. The study of [13] concentrated analysis on users and examined the changes of user profiles. The above works mostly used traditional charts and studied spatio-temporal properties of PBS data, which didn't analyze multiple impact factors. Corcoran et al. [7] studied the influences of weather and calendar events on PBS usage pattern. However, they only gave out the analysis results simply, and the users were not able to explore expected information satisfying their needs interactively.

In order to explore different aspects of data attributes, multiple linked-views are designed [14,15]. Drawing arrows or markers on map [14,16] is the most frequently used method to demonstrate the geographical context. As the number of flow increases, the visual result become cluttered. To overcome the occlusion problem and offer a more abstract depiction, several alternative methods have been proposed, such as edge bundling [17], OD matrix [18] and circular layouts [19,20], etc. Simple OD matrix visualization [18] has difficulty in recognizing paths due to the lack of spatial information. Circular layouts is able to display geographical entities' relationship and support including more information by using the peripheral rings. For temporal visualization, timeline [21] and table-like visual metaphor [22] are popularly used. Space time cube [23] draws the time as z-axis perpendicular to map, which shows temporal information along with the spatial context, but it might be hard to decode the information and make comparison. In this paper, we combine map and circular layouts together to show spatial attributes and adopt a table-like design to visualize temporal attribute, which could be understood by ordinary users easily.

2.2. The visualization of multi-dimensional attributes

There exists many techniques to visualize multi-dimensional data, such as polar charts, pixel bar charts [24], table lens [25] and parallel coordinates [8,9], etc. Among which parallel coordinates is one of the most widely applied methods of multi-dimensional visualization. It draws N parallel and vertical axes side by side. A point in n-dimensional space is represented by a polyline with vertices crossing the parallel axes. Data summarization approaches [26,27] were introduced to reduce line over-plotting.

However, parallel coordinates was used for displaying continuous variables, but not a good representation for categorical variables. The discrete nature of categorical data made it unsuitable for using traditional parallel coordinates [28]. Kosara et al. [29] presented "Parallel sets" to exhibit multi-dimensional categorical data variables. Von et al. [30] dynamically linked a map display with the space-time cube to enable visual exploration of spatio-temporal categorical data. VisBricks concept [31] further extended the idea of Parallel Sets to demonstrate the topology of large and inhomogeneous dataset. In this paper, we propose a modified parallel coordinates representation to help users analyze continuous and categorical attributes interactively in a unified view.

3. System design

3.1. System overview

An overview of our system is illustrated in Fig. 1. The required data, including PBS flow data, climate data and calendar data, are collected and stored in the database, as the input of visual analytic views.

We obtain PBS data in Hangzhou from April 2014 to June 2014. The key information is stored in two tables: *leaseInfo* and *stationInfo*. Table *leaseInfo* stores all records of bicycle loan, whose main attributes include *bikeID*, *leaseTime*, *leaseStation*, *returnTime* and *returnStation*. Table *stationinfo* stores the ID, address, longitude and latitude of a station. Because the original data has noise, a data cleaning process is adopted.