Analysis and Visualization of Public Transport for Integrated monitoring dashboard: Case Study of Thane, Maharashtra, India

Madhuri Patel
Department of Computer Engineering
Pandit Deendayal Energy University
Gandhinagar, India
Department of Information Technology,
L D College of Engineering,
Ahmedabad, India.

Samir B. Patel
Department of Computer Engineering
Pandit Deendayal Energy University
Gandhinagar, India

Debabrata Swain
Department of Computer Engineering
Pandit Deendayal Energy University
Gandhinagar, India

Parth Raval
School of Technology,
Pandit Deendayal Energy University
Gandhinagar, India
Department of Science & Humanities,
L D College of Engineering,
Ahmedabad, India.

Received: June 24, 2021. Revised: June 13, 2022. Accepted: July 21, 2022. Published: August 4, 2022.

Abstract—Urban population of India and migration towards large cities has been increased in last decade. Correspondingly civic management has to deliver various facilities on timeline based for several services. Smart City mission launched by Government of India focuses on several attributes where Intelligent Public Transport System is major focus. Currently 68 major cities in India have public transport system in city. 83 of the 100 shortlist smart cities have implemented an Integrated Command and Control Center (ICCC) to gather, analyze, evaluate, and respond to big data inputs from across the urban area. We have reported various critical attributes for visualization and analytics of Intelligent Public Transport system. Real database of Thane smart have been taken for academic research purpose. Various analysis reports have been prepared and correspondingly visualization for dashboard have been successfully created. All reports contain interactive details with attributes including 3D trip layers. Common data tables designed for various analysis have also been suggested. It is recommended to have common features for critical attributes for Intelligent Transport System (ITS) for all smart cities for better common implementation policy including fleet size, and shifting to Electric Vehicles (EV).

Keywords—Intelligent Transport System (ITS), Visualizations, Big data analysis

I. INTRODUCTION

The Urban population of Indian has grown in significant way first time from 2001-11. In year of 2020, 33% of the total population in India lived in cities which is 4% increment in Urbanization. The National Commission on Population (NCP) in India predicts that in the next 15 years (i.e., by 2036), about 38.6 percent of Indians (600 million) will live in urban areas. The UN, too, highlights that India’s urban population size will nearly double between 2018 and 2050, from 461 to 877 million. However, the level of urbanization in India is low as compared to the BRICS economies such as Brazil (86.6%), South Africa (66.4%) and China (59.2%) (WUP 2018). The level of urbanization across states is asymmetric. As per 2011 census, Tamil Nadu (48.4%) was the most urbanized among major states, followed by Kerala (47.7%), Maharashtra (45.2%) and Gujarat (42.6%). These four states together contributed to around one-third of the total urban population of India.

In view of the same, India has launched the mission being implemented under Smart Cities which has three core features as Liveability, Economic ability and Sustainability. Aggregated at the national level, the 100 Smart Cities have proposed to execute 5,151 projects worth INR 2,05,018 crores in 5 years from their respective dates of selection. Financial innovation is built into the capital investment plans. The distribution of funding envisaged from different sources is approximately as follows - 45 percent from central and state governments (INR 91,000 crore), 21 percent from convergence with other programs (INR 42,000 crore), 21 percent from Public Private Partnerships (INR 42,000 crore), 4 percent from loans (INR 10,000 crore), and 9 percent from own and other sources (INR 20,000 crore). The largest area of projects was sanctioned in field of Public Transport System. Currently, India has public transport systems in 84 cities out of which 57 cities have Bus Rapid Transport System as well.

Given the importance of Intelligent Transport System (ITS), many researchers have already contributed in area of traffic management and transport planning. In general, these researches need to possess an accurate picture of the underlying movement patterns [1]. However, the majority of empirical transport researches on understanding human movements rely on travel surveys [2] that have obvious limitations, including high expense, small samples and low update frequency [3,4]. In last decades, development of sensing devices like GPS devices, laser scanners and video recorders, massive amounts of urban public transport data have been collected in an automatic and pervasive way, providing transport researchers a superior alternative to study the movement patterns. The Smart Cities implementation task...
force has emphasized on data collection, data analytics and Integrated Command and Control Centers for monitoring of these data.

The automated mining of necessary information from the data, numerous data analysis methods such as data mining and statistics can be employed. It is observed that these methods still have substantial challenges such as algorithm scalability and data heterogeneity [5]. More importantly, these approaches have the problem of understanding and analyzing the analyses: The results are only reliable in well-defined and well-understood problems [6,7]. Visual analytics is an interdisciplinary research field that combines various related research domains such as visualization, data mining, human-computer interaction, and application domains, such as transport and GIS. Existing procedures and models in application domains can contribute to the development of effective visual analytics tools. Data science analytics have been applied widely to transportation data. For example, the urban data mining algorithms to accurately classify ground transportation modes (e.g., bus, car, bike, or walk) of commuters by analysing their GPS data accelerometer data and/or dwell time history have been successfully reported [8-9].

Predictive analytics have been conducted on various real-life situations outside the scope of transportation [8-14]. For those works that dealt with transportation data, many focused on predicting bus arrival times. For instance, Sun et al. [15] predicted the bus arrival time based on a geographic information system (GIS)-based map-matching algorithm. Lin et al. [16], on the other hand, employed GPS data and automatic fare collection (AFC) system data to predict bus arrival times through the use of artificial neural networks (ANN). However, auxiliary information like GIS and/or AFC system data may not be easily accessible to the general public. Instead, Rajput et al. [17] examined New York City open data, which are accessible to the general public. By applying clustering to the New York open data, they (a) identified areas with high congestion and insufficient bus stops and (b) recommended the installation of new bus stops in those areas. The most closely related work is our previous work [18] on an intelligent system for predicting the severity of bus delay.

However, there is very little amount of work to develop visual analytics for transport researchers to explore and analyze massive urban public transport data. It is mainly due to special focused issues like: (a) big data volume: the data consists of millions of commuter trips, bus trips in each day; (b) complex network: the public transport includes around hundred bus stations and thousands bus stops together in trips; (c) spot adjustments of trips: the traffic managers decides changes in bus in another routes and trip truncation and extension; and 4) spatial and temporal patterns: the movement patterns can vary drastically in both time and space dimensions.

The overall objective is to develop effective visual analytics systems for transport researchers to facilitate the understanding of overall public transport system, their optimization, real route analysis from the massive urban public transport data of Thane Municipal Transport System. We have meticulously track the agenda in the information visualization reference model [19]. This model describes the process of developing interactive visualizations motivated by a user’s task in three steps: (a) transform the raw data into data tables; (b) map the data tables into visual structures; and (c) build visualizations from the visual structures. The user can observe the views, and interactively control the process by adjusting any of the steps, if necessary. The detail analysis has been carried out for TMT transport system for more than 80,000 trips per month and more than 4GB tracking and Ticketing data of the same. The data management, hard ware data log system tables, dataset for the proper visualization has been suggested. The same will be extremely helpful for upcoming implementation of Integrated Command and Control System across all smart cities caters to Intelligent Transport System (ITS).

I. II. METHODOLOGY

Thane Municipal Transport System is spread across large area having 158 routes including special routes, more than 460 operating buses, 80,000 trips per months and 32 Lakh average tickets. Raw database is recorded in three different ways: (a) Live GPS data from bus caters live location (b) E-ticketing system data punch by conductors (c) Trip allocation data with route recorded at depot. During the inception of smart city projects, most of public transport services were completely executed by local civic authorities. The transformation of system to ITS were done in phase manner with different aspects. The complete coordination and integration are difficult task due to various implementation agencies and different timelines of implementation. It is definitely expected to have minimum standards that require transit agencies to use various analytic methods. It helps the civic bodies and funding agencies to estimate the actual implementation, future requirements and suggestive outcomes on day to day operations. Improved visualizations and analytic reports are useful for valuable transit data to provide a wide range of insights for transit agencies. It is from whether buses are arriving on time to which bus stations are used most or revenue collections with supply ratios that improves transport planning.

The current work emphasis on creation of various user-friendly reports and day-to-day analysis of TMT ITS data. The reports are divided into various segments such as Speed violation reports, skipped stops reports, Missed trips reports, Improper stopping reports, Route deviation reports, Harsh breaking/rapid acceleration reports, Schedule adherence reports, ticket collection reports, mis punch of routes reports, truncated journey reports etc.

II. III. REQUIREMENTS AND DATA DESCRIPTIONS

Services of Thane Municipal Transport (TMT) is spread across Bruhad Mumbai Mahanagarpalika. The city of Thane is located on the North-East of the salsette island and on the Northern extremity of Greater Mumbai. The boundaries of Greater Mumbai and Thane are contiguous. The city falls on longitude 72.50' North with longitude 19.10' East. According to the 2011 census, the area of divided district is 4214 sq. Km. Thane city has population of 18,41,488 and its metropolitan population is 1,83,94,912. The major transportation used by citizen is through local trains for far distance travel. The stations are internally connected by public transport stations where small-scale commute is possible. However, the large routes are also existing. The Public Transport system consists 459 operational buses, total 158 routes covering 721 stops, and the total transit routes cover 2182 km.

In this study, the visualization analytics was done through local transport transit details, GPS tracking data and trips data entered at depot. We analyzed the current practices of transit
agencies with regard to translating large volume transit related data into insights. Based on the same, the following potential analysis are (a) Route Analysis (b) Station wise analysis (c) Ticket collection analysis and (d) Monitoring transit operation and performance. The specifics of this scenario include analytics that automate the characterization of routes similarities, identifying crowded stations as a result of new scheduling plan (service changes), and so on. It is observed that available transit data has grown exponentially over the period of time. We follow an approach which leads to a development of an interactive visualization tool that fits with realities and needs. This study aims at filling in this gap by introducing a development process of a public transit integrated dashboard that measures and dynamically visualizes public transit operation from six perspectives (visualization modules); mobility, speed, flow, density, and headway as well as analysis module that summarizes and shows insightful analytical information.

This study is carried out using two basic tables as an input where the dynamic GPS data of each bus gives certain details which is described table (1). The trip wise manual data entered by bus conductor at time of departure is mentioned in table (2). Based on these raw data, the various reference tables were prepared like (a) Route Master (b) Waypoint Master (c) Trip Waypoint details with distance (d) Geolocation files and (e) trip master. The developed visualization tool is based on the transport service data between March 2022 to May 2022.

In addition, the Geographic Information System (GIS) shapefile of Mumbai Metropolitan Region is also used. This enables to give perfect visualizations of entire system.

**A. Data preprocessing**

The raw data is required to be preprocessed for accurate visualizations and analytics. There were mainly three steps in our data preprocessing that includes data cleaning, reorganization, extraction, and filtering (Figure 3).

**B. Raw data**

Our raw data from TMT was 4GB MB in total per month, consisting of trip data, tracking log data, Ticket data table in database files.

**C. Data cleaning**

It was expected about errors in the raw data like incorrect and missing records. This is mainly due to GSM signal strength, manual data entry mistakes, immediate reallocation of routes and truncated routes. Some of the records were due to incomplete attributes like missing geographical location records of waypoints which we removed from our database. Some of missing location between a trip for which we interpolated the lost value based on the head-to-head locations.

**D. Data reorganization**

The integration of various data and its corresponding structure files has been prepared for easy access for analytics and visualization. The major focus was to obtain actual details
of actual trip conducted, trip time, waypoint wise trip time, number of tickets sold, and other operational details.

E. Analytics preparation

After data verification and reorganization, different perspectives of the transit system operation modules are prepared like route wise, time wise, station wise, waypoint wise, mobility, speed, flow, density, headway, and analysis. Various analysis has been prepared which gives clear idea about entire transport detail.

F. Visualization

The visualization consists of user friendly, easy trend identification, and interaction based graphical outputs. For the same, system visualizes the transport network as whole in which various analysis have been carried out like: day wise-route wise, weekday/weekend wise-route wise, station wise total actual departed routes, time wise route-wise segments, break down or delay transit operation, ideal time of bus, waypoint wise stop time and start time. The entire transport network is been shown in Fig. (). The time stamp trip layers suggest us possibilities of the route-wise restructuring and circular routes.

The TMT routes are almost running more than 20 hours everyday. The early morning route starts from 4.00AM in morning and last trip takes at 1.00AM in late night. Fig() shows timewise route dispatch from every station. The details show apart of Thane SATIS station, the other significant stations can also be identified. This helps for mobilization of resource and allied staff. The major trips occurs in morning and evening times. On the similar ground, fig() shows time slot wise dispatch of number of trips. In recent times, the involvement of Public Private Partnership and route-based contract system has been introduced in various smart cities.

Above analysis helps for better allocation of resources, optimum usage of buses and shifting of the buses in emergency. In migration of traditional diesel-based buses to Electric Vehicle (EV) bus system the time-route-station wise allocation helps to migrate gradually. This also enables them to allocate or deploy small size buses where utilization is limited.
Identification of waypoints where start and stop time have been done. Detail bus wise analysis of idle time suggests that after 10.30 AM, idle time increases to one hour sometime. Actual stoppage report has been given in table (1). Bus can be utilized or shifted accordingly between 9.30AM to 1.30 PM.

<table>
<thead>
<tr>
<th>Time Stopped</th>
<th>Total Minutes (hh:mm)</th>
<th>Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>05:59:50 TO 06:10:46</td>
<td>00:11</td>
<td>Road Number 34, Wagle Industrial Estate</td>
</tr>
<tr>
<td>07:22:13 TO 07:28:38</td>
<td>00:06</td>
<td>Rabodi</td>
</tr>
<tr>
<td>08:04:23 TO 08:15:38</td>
<td>00:11</td>
<td>Road Number 34, Wagle Industrial Estate</td>
</tr>
<tr>
<td>08:41:40 TO 08:48:10</td>
<td>00:07</td>
<td>Pant Ram Marathe Path, Rabodi,</td>
</tr>
<tr>
<td>09:14:52 TO 09:51:01</td>
<td>01:03</td>
<td>Road Number 34, Wagle Industrial Estate</td>
</tr>
<tr>
<td>10:31:50 TO 10:57:44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:58:35 TO 11:59:49;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12:00:45 TO 13:16:48</td>
<td>02:17</td>
<td>Road Number 34, Wagle Industrial Estate</td>
</tr>
<tr>
<td>13:16:58 TO 13:21:09</td>
<td>00:05</td>
<td>Road Number 34, Wagle Industrial Estate</td>
</tr>
<tr>
<td>13:40:05 TO 13:50:42</td>
<td>00:10</td>
<td>Vartak Nagar, Thane</td>
</tr>
<tr>
<td>14:09:33 TO 14:17:43</td>
<td>00:08</td>
<td>medos, Station Road, Rabodi</td>
</tr>
<tr>
<td>14:37:14 TO 14:52:08</td>
<td>00:15</td>
<td>Vartak Nagar, Thane</td>
</tr>
<tr>
<td>15:11:49 TO 15:20:20</td>
<td>00:09</td>
<td>Thane Station, Station Road, Rabodi</td>
</tr>
<tr>
<td>15:36:56 TO 15:58:41;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16:38:24 TO 17:30:37</td>
<td>01:14</td>
<td>Vartak Nagar, Thane</td>
</tr>
<tr>
<td>17:52:48 TO 18:04:24</td>
<td>00:12</td>
<td>Thane Station, Station Road, Rabodi</td>
</tr>
</tbody>
</table>
This time can also allot to EV charging schedules when migration of complete electric vehicle is done. Similar way the other important analysis tools have been developed with visualization like (a) speed chart bus wise (b) vehicle route reports – bus wise (c) acceleration report (d) hardbreaking report with driver i’d (e) route wise turnout report with ticket collections etc. Such analysis and its integrated dashboard are extremely important for monitoring and policy implementation purpose. This helps in optimum utilization of resources, clear understanding of fleet health and migration towards new technologies. It is important to note that the large size data requires non-traditional methods to execute the desire system.

III. IV. CONCLUSION

Public transport system is essential backbone for smart cities. Smart mobilities and ease of access of public transport is essential part of growth of city. India has large growth towards urban population and migration towards large cities in last decades. The migration from conventional operations to decision based on data visualization is required. The integrated command and control systems for smart cities will provide the robust support for data analytics and visualization. The common tables and common database will help to integrate entire systems for all smart cities especially upcoming smart city infrastructures. During the integration of data exact routes, truncated routes, day wise slot wise routes, route-wise time-wise ticket collections, optimum route utilization, idle times can be easily visualized. It is recommended to finalized minimal features of some crucial attributes by Niti Ayog, Government of India for integral analysis of smart cities. The restructuring of routes majorly focuses on circular path is possible. This will enhance the utilization of resources. The small fleet buses in electric vehicle fleet is also suggested. We have successfully implemented real-time data based visual analytics modules like (i) schedule adherence analysis; (ii) headway regularity analysis; and (iii) speed, delay, and dwell time analysis.

IV. ACKNOWLEDGMENT

The authors are thankful to authorities of Thane Smart City for providing support, management of PDEU, Gandhinagar and LDCE, Ahmedabad for supporting research work. We are sincerely thankful to Mr. Siddharth Shah and Mr. Sanghani Yash for help in interface development. The authors are thankful to Mr. Munj Raval for suggested guidance for understanding the system.

REFERENCES


Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0
https://creativecommons.org/licenses/by/4.0/deed.en_US