

Chapter 9

Study and Analysis of Delay Factors of Delhi Metro Using Data Sciences and Social Media: Automatic Delay Prediction System for Delhi Metro

Arun Solanki

Gautam Buddha University, India

Ela Kumar

Indira Gandhi Delhi Technical University for Women, India

ABSTRACT

Delhi Metro passengers had a difficult time mostly on Monday morning as trains on the busy corridors are delayed due to technical problems or track circuit failure. This study found different factors like power failure, weather, rider load, festive season, etc. which are responsible for the delay of Delhi Metro. Due to these factors, Metro got delayed and run at a reduced speed causing much inconvenience to the people, who are hoping to reach their offices on time. Delhi Metro data are received from different sources which may be structured (timings, speed, traffic), semi-structured (images and video) and unstructured (maintenance records) form. So, there is heterogeneity in data. Except for this data, the feedback or suggestion of a rider is vital to the system. Nowadays riders are using social media like Facebook and Twitter very frequently. Three-tier architecture is proposed for the delay analysis of Delhi Metro. Different implementation techniques are studied and proposed for the social media module and delay prediction modules for the proposed system.

INTRODUCTION

The Delhi Metro, an intra-city electric rail system serving the National Capital Region (NCR), has been operational since December 2002. Metro services did not stop entirely due to the faults; it delayed for more than an average time. It has also watched in working hours that Metro is not on its schedule. As,

DOI: 10.4018/978-1-5225-3176-0.ch009

Delhi Metro trains run on automatic signaling system, and if a problem occurs, it operates on a manual signaling system. So, there is an acute need for study and analysis of this problem and to develop a system as a solution. The related data is collected from different types of databases like the Relational database, Transactional database, Multimedia database, spatial database, Time-series database, World Wide Web, etc.

The riders are the most important part of the Metro, and Whenever a passenger stuck in the metro due to delay, rider updated a Facebook status or tweets about the issue. A real-time data is generated and may be captured for analysis the problem. As this data may be big data; Hadoop or Map Reduce, which are advanced tools of data science used for sentimental analysis of tweets or Facebook updates. Data mining is done on both types of data i.e. operational data and rider social media data. The first and essential step of data mining is preprocessing which is applied for data cleaning, data integration, data transformation and data reduction. These steps of preprocessing filter the data from errors and noise and further used. This architecture has Data Acquisition, Data Processing, and Disruption Analysis & Results components. Data acquisition contains the data received from social media and operational data of Delhi Metro. These data may be in structured or unstructured form. Data Processing has three modules named Master Data, Integrated Data and Internal Analytics Aggregation. Disruption Analysis & Results contains the trends, application, and impact of disruption of Delhi Metro.

To handle big data problem Data science is used. Data science is also known as data-driven science which is an interdisciplinary field the scientific methods, processes and systems to extract knowledge or insights from data in various forms, either structured or unstructured, similar to Knowledge Discovery in Databases (KDD). Ali et al. (2016) discuss in their work that Data Science looks to create models that capture the underlying patterns of complex systems and codify those models into working applications while Big Data seems to collect and manage large amounts of various data to serve large-scale web applications and vast sensor networks. There are more than 40 data science techniques like clustering, search engine, deep learning, neural networks, Hadoop, etc. which can be used for analysis of delay factors which impacts the Delhi Metro. A predictive model is generated using multiple regressions to predict delay in Delhi Metro. In next section delay factors which mainly affect the Delhi Metro are discussed

DELAY FACTORS IN DELHI METRO

Identification of the delay factors and their reasons helps the future planning of service, and it can be used for traffic forecast. The delay factors can be broadly divided into two parts:

1. **Operational Side:** Fan and Weston (2012) discuss operational side algorithms like brute force, first-come-first-served, Tabu search, simulated annealing, genetic algorithms, ant colony optimization, dynamic programming and decision tree based elimination are already examined
2. **Passenger Side:** Nagy and Csiszar (2014) in their work discuss the specified factors (weather conditions, lines, service type, etc.) and the punctuality of vehicles (departure and arrival time) is predictable. These values can be used for passenger information on stations as well as on personalized travel information applications like journey planners.

As a rider, it has been watched many times that Delhi Metro is not on the schedule. Factors like power, weather, rider load, network congestion, festive seasons, speed and others are responsible for the

13 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/study-and-analysis-of-delay-factors-of-delhi-metro-using-data-sciences-and-social-media/191747

Related Content

Analysis of Heart Disease Using Parallel and Sequential Ensemble Methods With Feature Selection Techniques: Heart Disease Prediction

Dhyan Chandra Yadav and Saurabh Pal (2021). *International Journal of Big Data and Analytics in Healthcare* (pp. 40-56).

www.irma-international.org/article/analysis-of-heart-disease-using-parallel-and-sequential-ensemble-methods-with-feature-selection-techniques/268417

A Markov-Chain-Based Model for Group Message Distribution in Connected Networks

Peter Bajorski and Michael Kurdziel (2020). *International Journal of Data Analytics* (pp. 13-29).

www.irma-international.org/article/a-markov-chain-based-model-for-group-message-distribution-in-connected-networks/258918

Tools, Technologies, and Methodologies to Support Data Science: Support Technologies for Data Science

Ricardo A. Barrera-Cámara, Ana Canepa-Saenz, Jorge A. Ruiz-Vanoye, Alejandro Fuentes-Penna, Miguel Ángel Ruiz-Jaimes and Maria Beatriz Bernábe-Loranca (2021). *Handbook of Research on Engineering, Business, and Healthcare Applications of Data Science and Analytics* (pp. 50-72).

www.irma-international.org/chapter/tools-technologies-and-methodologies-to-support-data-science/264304

What Is Open Source Software (OSS) and What Is Big Data?

Richard S. Segall (2022). *Research Anthology on Big Data Analytics, Architectures, and Applications* (pp. 77-118).

www.irma-international.org/chapter/what-is-open-source-software-oss-and-what-is-big-data/290978

Patent Information Quality to Stimulate Innovations

Sérgio Maravilhas, Sérgio Goes Oliveira and Paulo Melo (2019). *Handbook of Research on Expanding Business Opportunities With Information Systems and Analytics* (pp. 120-144).

www.irma-international.org/chapter/patent-information-quality-to-stimulate-innovations/208561