ROUTED – Dynamic Bus Scheduling using Machine Learning

By

Rushabh Kapadia	141080014
Dharmit Prajapati	141080038
Kevin Daftary	141080055

Under the guidance of

Prof. Mahesh Shirole

Computer Engineering and Information Technology Department

Veermata Jijabai Technological Institute

2017-2018

TABLE OF CONTENTS

Chapter 1

Project Title: “Routed”

INTRODUCTION:

In this day and age of daily commute, public transport as a service has a key role to play. Bus transportation is the primary mode of transport by road for most people. There are on an average 8.5 lakh people commuting by buses spread across 135 bus routes in Island city, Mumbai. While most people have to travel in jam-packed conditions to get their day started, there are empty buses running on the routes they take as well as on the routes not often travelled by commuters, which is a waste of resource that is needed on some different routes.

This is mainly because of the inefficient bus routines and high inter-arrival time between buses on the same route that is currently observed as a by-product of the static mundane bus schedules, because of which people cannot afford to miss their daily favoured bus that gets them to their workplace on time. 75% of the Island city is connected by high frequency buses and yet, the bus services suffer from overcrowding.

The growth in computer technology can come to the rescue by helping the service identify the number of buses and their schedules on the specific routes allotted to them. Such a proactive system would help ease the hassle-filled experience of daily commuters, which would help build a sense of trust and responsibility towards the public transport services among the masses. Moreover, by ensuring dynamic schedules tailored for the specific routes based on the frequency of people commuting along the route would ensure better utilization of the buses on the route as well as efficient fuel consumption and utility, ensuring better experiences for commuters as well as profitability for the service providers.

THE PROBLEM:

Because a public transit agency is an unusual hybrid of private business and government, transit planning is a difficult profession. Should transit focus on providing transportation for those with no other choice, or should it strive to become a competitive alternative for the car? Unfortunately, it is difficult to simultaneously serve both alternatives. This difficulty is frequently aggravated by political interference in the transit planning process, which often forces transit agencies to operate inefficient bus routes and to construct sub-optimum rapid transit projects.

As a result, the existing BEST bus services use a static and inadequate system of routing buses on various routes which is quite inefficient from the economic point of view and overutilization of resources. The current Bus schedules are static and fixed which do not vary the frequency of buses along the routes according to the frequency of travellers at specific times of the day, which leads to overcrowding of buses. The current Bus schedules also have a fixed number of buses running on each route which leads to many buses running along routes without being utilized by passengers at all due to the time at which they are travelling along the routes and because the buses are being driven on routes that already have enough buses to cater the needs along the route.

A project to help the Bus service to better utilize the buses has been developed but it only considers routes in one direction as well as just helps to predict how many buses are needed along each route. Another shortcoming of this solution is that it considers unlimited number of buses being available at the disposal of BEST.

EXISTING SOLUTION:

The existing solution used by the public transport agency relies primarily on human schedulers who decide the routes travelled and the stops to be kept on the given route depending on the traffic analysis and by considering the demographics of regions such as commercial areas, residential areas, areas where traffic is high. The existing solution decides and finalizes the schedules by manually travelling along specific routes and finding out the frequency of people travelling along the given route at different times of the day. This approach is quite

tedious since it requires a large human workforce to exhaustively cover all the routes and travel along these routes at all possible times of the day.

This solution also requires a very drawn out and lengthy period of research before the bus schedules are even created and finalized. As a result, this approach usually takes 2 to 3 years of research before even considering whether the bus schedules need to be changed. This imposes an inherent inability on the existing system that ensures that the bus schedules cannot be dynamic and redesigned on monthly basis, let alone weekly basis.

However, the benefits of this approach are the fact that it ensures that there is no ambiguity or a factor of unpredictability of bus schedules since the data considered has been obtained through thorough observation. It also ensures that factors such as construction of new roads or construction of new skywalks, etc. along the existing routes are considered beforehand and can be accounted for in the designing of bus schedules and distribution of buses across different areas and routes.

THE PROPOSAL:

The solution proposed here, mainly focuses on improving and better utilisation of Buses along various routes with dynamic schedules being generated to prevent overcrowding o buses during peak hours. The solution has mainly been divided into two modules for decoupling of two separate processes being carried out –

1. **Number of Buses Prediction:** Predicting the optimum number of buses required along each route by analysing the dataset from the previous week using clustering algorithms and regression models for prediction.
2. **Schedule Generation:** Once we have the optimum number of buses required along each route, the next process is to schedule these buses optimally along the routes such that it ensures better overall experience for passengers and prevents overcrowding of buses during peak hours. This will also ensure that during peak hours, the amount of waiting time for passengers for a particular bus would be restricted to a minimal value.

This solution aims to create dynamic schedules for BEST bus services to ensure better utilization of resources (buses, fuel, etc.) and cutting down unnecessary losses due to incumbent static schedules currently being used.
The solution will take into consideration constraints like:

- Fixed number of buses to be optimally distributed across various routes
- Reusable buses i.e. same bus can travel along the given route multiple times in both directions
- Passenger centric schedules, i.e. considering the frequency of passengers at different time slots and accordingly varying the frequency of buses running along the route in that timeslot
- Only a single time of bus with an arbitrary capacity has been considered rather than many different types of buses that are available.
- The route is bi-directional i.e. the buses can travel on the same route in both directions.

The solution will not be considering following cases/instances like:

- Overlapping routes i.e. routes which have some of their part overlapping with each other
- Dynamic allocation of buses along the routes within a day based on timeslots i.e. once the buses have been distributed along the routes for the given day, they will not be re-allocated to some other route during that day
- Economic factors such as trying to reduce the cost of travel for passengers by running the same bus on a combination of two or more routes.
- It will only consider pre-existing routes from the start stop to the end stop i.e. the buses once scheduled will travel the whole route from start stop to end stop and not just a specific subset of the route.

Chapter 2
Literature Survey

In our literature survey regarding the relevant domains and research papers to be focused on for the proposed solution, we mainly focused on research work in the fields of Cluster Algorithms, Bus Scheduling algorithms, optimal resource allocation, etc. Our prime focus during the literature survey has been to find out research papers from various sources that can be realized or cited in our solution.

The primary hurdles while performing the literature survey has been finding relevant research in alignment with our project objectives and that can be modelled or adapted to a real-time bus allocation and timetabling environment. The approaches considered within these research papers vary widely from statistical methods to data mining models and linear programming for Bus scheduling. Another type of research work considered has been identifying the trade-offs between different attributes/features that can affect the creation of optimal Bus Scheduling. Papers also considering the prediction of continuous values have been surveyed to identify predictive algorithms that can be adapted for the given problem of Bus schedule prediction.

With respect to Bus/resource allocation approaches, our primary aim has been to identify different types of research papers, such as papers proposing mathematical models for Bus Allocation, papers defining various angles to approach the problem of Bus Allocation and papers focusing on the identification of different attributes/features in the data that can be exploited for resource allocation have been surveyed in detail.

Another prime area of focus has been on identifying suitable clustering algorithms and their adaptations to problems similar to the Clustering problem in our proposed solution. Papers on modified K-means and various other clustering algorithms/aggregation methods have been surveyed for their suitability to the current discussion.

Also, as an additional reference point, some of the already existing bus scheduling and allocation models proposed previously have been studied to identify and concisely define the problem that we wish to target. A few Data-cleaning approaches and their implementation considerations have also been surveyed to support the data cleansing pre-processing that might be required.

The research papers considered are in no way the most exhaustive ones but they mainly cover all the domain areas that we wish to cover and implement in our solution. Also, another important factor is the fact that these papers have been drawn from a restricted pool of free research papers that are easily accessible on the internet, thus, excluding the paid research papers. All in all, the literature survey provides us with a basic foundation and an understanding of the problems that we might encounter during implementation and the difficulties associated with each approach.

Our main take away from this survey is to understand in depth the algorithms and approaches we might consider but in no way are bound to implement. The research papers have mainly been studied as being a guiding direction and not as something that is to be implemented and compared with the theoretical results.

We have attached the summary of the research papers that we have surveyed below.