International Journal of Advanced Research in

Computer Science Engineering and Information Technology

Volume: 5 Issue: 3 22-Mar-2015,ISSN_NO: 2321-3337

A TECHNOLOGICAL FIX FOR SUSTAINABLE URBAN TRANSPORTATION

Mr. Kedar R. Joshi

University of Pune

ABSTRACT- *the Public Transport Systems run by local authority of state government has deficiency in designing of bus schedules. Most of the times the traffic patterns observed are inconsistent. There are no. of factors related to this inconsistency. This research study will analyze the bus-route scheduling, frequency distribution in peak hours, shift towards private motorized, non-motorized transportation with the help of association rules applicable to PMPML (Pune Mahan agar Parivahan Mahamandal Limited) under the working boundary of Swargate Depot.*

KEYWORDS: Public Transport Systems issues, planning categories, PMPML (Public Transport-Swargate Depot.)Case, PS monitoring, association rules, inconsistent traffic patterns.

1.INTRODUCTION

Public transportation is a field that is generally well suited to the application of mathematical optimization methods. In many municipal corporations public transport is organized in a centralized, completely scheduled way, time and theoretical data for planning is available and lastly a number of important planning problems are close to classical optimization problems.

In fact, the influence goes also the other way round, i.e., scheduling problems in public transit have been among the driving forces behind the development of several areas of applied mathematics such as discrete and combinatorial optimization and integer programming. The Term Regularity is the key for data mining for assessing any public transport. If we consider any give bus route as a sample then that bus will run at given frequency, on allocated route, by attending every Bus stop, so this characteristic will help in collecting information regarding the traffic patterns.

We have to take into account different parameters associated with Road (length, speed restrictions, terrain geometry etc.) and the daily cycle of buses, predicting other schedule of a public transportation system should be possible. But often the variations in traffic make it very difficult to determine an acceptable approximation to the arrival and departure times which will also leads to administrative issues such as over time. The Traffic patterns and traffic flow are unpredictable at Rush hours, weekends or holidays. Also some extra things along a bus route - like traffic lights and pedestrian crossings - represent for the driver elements of randomness which can be foreseen only with a certain probability. Hence a reliable integrated prediction system is needed, one that should take into account both the relationship between traffic events and date & time information, as well as the probability of random events like stopping at the traffic lights. In this paper the case of the PMPML particularly of Swargate Depot is considered for - total no. of different routes (53 in no.) [1] , The total no. of Buses available in depot (170 own + 70 on rent) [1] .

The classification of buses depending on different criteria applied such as 1) On the basis of type of Fuel Required Diesel or CNG, 2) [1] On the basis of Size of Bus Large, Medium, Small, types of buses (own-10 in no. fuel-Diesel size-Large ,fuel-CNG size-Medium ,on rent-18 in no. fuel-Diesel size-Large , 52 in no. Fuel-CNG Size-large) [1] , No. of Bus stops (18 within working Boundary), The Sources and Destinations of different bus routes, the total no. of Passengers traveled in a sample month (3249716)

The focus of research is on finding a solution for speed variations of buses, using data mining techniques on collected GPS information from public transportation and to investigate correlations between road features, date & time information, passenger count etc. speed, frequency distribution and design a schedule of buses at different routes by considering above mentioned aspects.

2.TECHNICAL COMPUTATIONS

In any public transport system generally following categories of operations are present:

1) Service Design: which includes

a)Network Design
b) Line Planning
c) Time Table Planning
d) Fare Planning

2) Operational Planning: which includes

a)Vehicle Scheduling
b) Duty Scheduling
c) Duty Rostering

3) Operation Control: which includes

a)Vehicle Dispatching

b) Crew Dispatching

c) Depot Management

We can map the public transport problems to the network optimization problems. When above data is considered then we came to know that there are no. of routes which has average or low frequency [1] but the total no. of People traveled on a given sample route is quite high so the frequency distribution is not even also the public transport suffers due to size of buses because many of the routes are going through the congested area. Ref data set for first 5+1 bus stops (22 stages, 66 bus stops for route No. 159P

So the existing system is indicating the above mentioned flaws which can overcome with some sort of feasible solution such as the authority can have the small size bus, if this factor is considered then that will represent a partial solution as we are not considering the total no. of people traveled on that route. Hence the frequency of these small buses should be kept moderate at non-peak hours and should increase to high for peak hours.

For the routes which going through the broader routes on that routes the large buses with moderate frequency should be kept for this purpose we can use the concept of Time slicing where a process can be considered as a bus on a route.Also in addition to that we can maintain the synchronization among the different routes collides at that particular route. Some additional measures includes designing of Circle Routes at average frequency.

The major hurdle in achieving the above said improvements is finance. Due to lack of finance most of the things are getting affected such as buying new buses, Maintenance charges, spare parts. To overcome this hurdle the advertisements on buses and on bus stops will play a crucial role. by using dynamic programming we can sort out the neural network generated by the given bus stop Also identify such bus stops where no. of buses passes per unit time is beyond threshold value there are other factors also such as location of bus stop, surrounding etc. Arrange the different stake holders for advertisements on the basis of payment by that stake holder.

On such bus try to maximize the function of revenue generation by having these constraints using simple method of Linear programming and we can have the precise value by using Optimal Binary Search tree. There is a Standard interface developed by the National Marine Electronics Association (NMEA) which includes GPS receivers and GPS GGA (Global Positioning System Fix Data).The NMEA contains the following information:

i) UTC timestamp in “hhmmss.sss” format (hh - hours, mm -minutes, ss - seconds, sss - miliseconds);

ii) latitude and longitude in “Dm,H” format (D - degrees, m - minutes with 4 decimals precision, H - hemisphere);

iii) GPS quality information: 0 for “invalid”, 1 for “GPS fix”, or 2 for “DGPS fix”;

iv) number of satellites being tracked;

v) horizontal dilution of the position: a measure of GPS accuracy, based on the geometry of tracked satellites; this attribute assumes a numerical value between 1 and 20, where “1” equals to the highest possible confidence level;

vi) altitude, followed by its unit of measure.

The Haversine formula is the one used for calculating geographical distances, more precisely - to compute the **shortest path** between two geographical coordinates on the surface of the globe (measured in degrees or radians). Let (ϕ_s, λ_s) and (ϕ_f, λ_f) be the (latitude, longitude) pair of two points s and f, measured in radians. Then the spherical (angular) distance of the two will be:

$$\Delta\hat{\sigma} = 2arcsin(\sqrt{sin^2(\frac{\Delta\phi}{2}) + cos\phi_s cos\phi_f sin^2(\frac{\Delta\lambda}{2})})$$

where $\Delta\phi = \phi_f - \phi_s$ and $\Delta\lambda = \lambda_f - \lambda_s$. The distance d (or arc length), for a sphere of radius r and $\Delta\hat{\sigma}$ becomes $d = r\Delta\hat{\sigma}$.

3.DATA COLLECTION MECHANISM

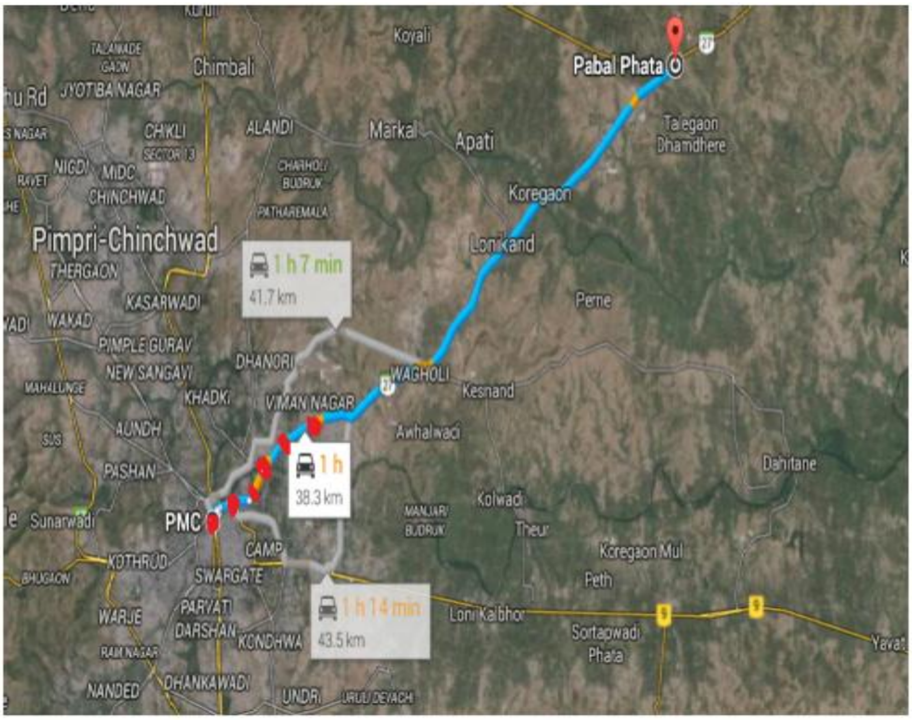
Initially, GPS devices are mounted on one or several vehicles along the examined bus route. The movement of a vehicle is going to be sampled by the GPS receiver at a given time rate, thus associating every piece of geographical information (latitude, longitude, altitude) with a timestamp. The devices should also offer the possibility to communicate with a remote server, task usually accomplished by GPRS, 3G or radio connections. As the GPS device records geographical information, it is sent via the remote connection to a central processing unit, which accomplishes storing, prepossessing and analyzing the data. Also by the means of this connection, the GPS device can be used not only for data mining but also for real time monitoring of the bus fleet. How long and how often we keep monitoring the activity on a certain bus line is a question of what kind of results we are actually expecting.

Traffic generally varies from workdays to weekends, and also according to the time of the day (e.g. rush hours) and to seasons (e.g. during rainy season or after rush/peak hours). For the latter case, our analysis will have to cover the whole year on a regular basis: monitoring a couple of days per week, for the whole year. Both for limited studies as well as in the case of yearlong surveys, The following rules should be considered:

i) the analysis should preferably cover the entire daily schedule on the given route; otherwise omitted time periods will be assigned to nearest available data and results for these gaps might be unreliable;

ii) there should be a balance between the number of samples taken on workdays and weekends;

iii) whether we provide GPS coverage for only one bus on the track or for all of them, at the end of the survey There should be enough data available to minimize the effect of outliers and technical errors; horizontal dilution can be used to filter unreliable information.



data set for first 5+1 bus stops (22 stages, 66 bus stops for route No. 159P

However, there are no studies of BRT systems in Indian cities that look into their accessibility by different social groups, especially poor urban communities. This We can study from the standpoints of sustainability and equity, based on the claims and promises in its two detailed project reports. The study has two guiding principles - sustainable mobility and equity. Sustainable mobility (prioritizing accessibility) describes all forms of transport that minimize fuel consumption and carbon emissions by minimizing the need to travel.

There is one other flow of thoughts that traditional urban transport planning based on the model and need of Mobility and on a thought or assumption that by increasing speed of traffic by reducing size of vehicles we can save travelling time but it is now leading towards the social, economic, and environmental problems. It is indirectly making it automobile (particularly Motorcycle) dependent which may result in "Lock-in" kind of transportation

3.1 BRT as a feasible solution

In the urban areas across India It has been seen a very high increase in the number of motorized vehicles in recent decades. The (NUTP) **National Urban Transport Policy** declared that "population increased 1.9 times during 1981-2001 of main cities, no. Of personal vehicles especially motor cycles increased by7.75 times." In India we have about 35% (about 27.76million) of urban households a Motorcycle two wheeler and 9.7% (about 7.65 million) four-wheeler. The rate of increase in recent years is 10.2% per annum. Even though this is the exponential growth but no. of vehicles per unit population is low i.e. 117 vehicles /1000people.

US (828), UK (544), Brazil (275) and South Africa (170). However, continued high economic growth will increase vehicle density in India in coming years. Besides different motor vehicles, urban traffic in India includes high levels of Non-motorized transport. According to Ministry of Urban Development as Public transport systems in different cities in country has not adapted to the situation of ever increasing demand for the public transport, hence that will compelling people to turn to either personalized modes or IPT (intermediate para-transit) which ultimately leads to the more traffic congestion and sound as well as air pollution. The local and State government has tried more option for increasing the sustainability in the Public transport through the road widening and flyover construction but they do not attempt to improve the Scheduling and conditions of buses itself

The National Urban Transport Policy (NUTP) identifies this root cause as a part of big project JNNURM(Jawaharlal Nehru National Urban Renewal Mission) and stated in their report as, There are no concrete methodology to identifies exact flaws and a remedial action plan as a result there are buses which has low/no maintenance, due to the lack of loose parts, lack of finance from local government, no sufficient human resources.

.

Though the fares of general public transport is kept minimum as a general public policy but there are no other measures to be implemented to increase the revenue, so it is very difficult to maintain each and every bus in a condition and have a routine maintenance, one solution can be thought of as identify the total no. of bus stops on given route R1 say nbs=66 then classify them in to different clusters by using K-means algorithm on the basis of different criteria (functions)such as " the amount of time spend on a given bus stop (where speed factor is 0)", No. Of. People entering and leaving the bus at that bus stop", after identifying such bus stops maximize that function for the revenue generation i.e. Apply more advertisements on these bus stops. AS far as the importance of public transport and equity concerned the NUTP has the following objectives. (i) To have more equitable allocation of road space among various users, more priority is given to the people than vehicles (ii) To encourage greater use of public transport and NMT modes by offering central financial assistance for this purpose; and (iii) To enable the establishment of multi-modal public transport systems that are well-integrated, providing seamless travel across modes Under the JNNURM's Urban Infrastructure and Governance (UIG) component,24.2% of the total allocations were for transport- related projects.

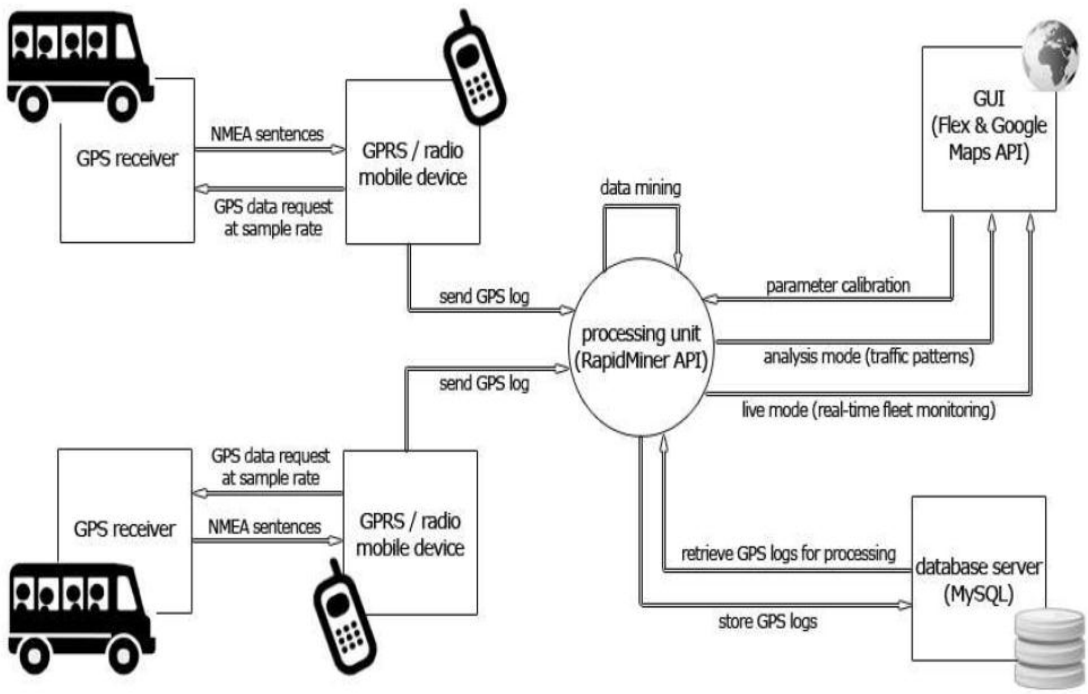
However, about 13.3% of this was for road widening and flyover building and only about 8.66% for mass transit, while the rest was for parking and other small transport projects. In spite of the NUTP's objectives, the largest portion of transport-related funds in the JNNURM was spent on roads. The NUTP supports BRT systems for their cost-effectiveness and many benefits as rail-based transport systems are too expensive to build and maintain in the long term.

In the Indian context, this can be shown by comparing the cost of the Delhi metro project with a BRTs. According to different reports by Government of India, The cost of the Delhi Metro Rail Corridor of about 190 kilometers (Phases I and II) is about Rs 29,702 crore (About $5.6 billion) or Rs 156 crore (about $29.5million) per km. For this particular amount, 2,900 km of a BRT network, at the cost of Rs 10 crore ($1.93 million) per km, could have been constructed .BRT systems have an advantage in the Indian context because of a multi-nuclei urban form, mixed land use, and trips distributed in multiple directions, also we can have different combinations of the categories of different vehicles such as Motorized vehicles which includes the Motorcycles and the Four wheelers also the non-motorized mode of transport which make average trip lengths low.

This kind of urban form and land use system also make it difficult to find high relational issues for efficient metro rail services.

4.GENERAL STRUCTURE OF THE SYSTEM

The following fig. depicts the architecture of system. On the left side we can have the data collection system, described on the previous section. The processing unit coordinates all data mining activities and is responsible for drawing the actual conclusions (rules) out of the knowledge base. Itimplements a Hadoop cluster-Nodes-Rack architecture or we can use its ancestor Rapid Miner methods through the Rapid Miner API.



A Feature Selection and Combination Data acquisition offered us valuable knowledge on the geographical coordinates (latitude, longitude, altitude) of the monitored vehicles, associating them to date & time references. We can also access data reliability using the horizontal dilution attribute. The finality of this research should be finding correlations (rules) between road features, time information, bus load etc. and the speed variations that actually stop us from assuming a fixed bus schedule. We shall combine the current knowledge in order to assess new features that would better explain the speed variations. By applying the Haversine formula on every two consecutive points we can obtain distance. Correlating this feature with the timestamp reference of consecutive points leads us to speed, which is a determinant attribute to our system.

Data information can be processed to obtain the day of the week or the season of the year. Altitude difference of two consecutive points could also be of help, as speed usually increases over the descent and drops when driving uphill. Another interesting feature that can be obtained from the raw GPS data sets is bus station interest. This feature automatically assesses passenger flow at bus stops and is deduced mainly from the time spent by buses loading and unloading passengers in every station. The more time a vehicle spends waiting at a particular station, the higher the passenger flow (in and out) and subsequently the higher the interest for this station, at the given timestamp. We thought about the crowding in downtown bus stops at rush hours compared to the flow in suburban bus stops on weekends. In order to automatically label a geographical point as being part of a bus stop we must assume a consistent volume of GPS information. For every log we process, speed is always going to drop to 0 around bus stations. Of course, this is also the case of road sections with traffic lights, pedestrian crossings, traffic jams etc. However, given a consistent knowledge base, we are able to make the following observation: whilst speed always levels to 0 around bus stops (given all stations are mandatory for the bus driver) for all other situations this will happen only by a certain probability (e.g. traffic lights and pedestrian crossings will generally have random effects, traffic jams only take place at rush hours).Thus, points which maintain a constant speed of 0 over all GPS logs will be labeled as bus stations and we may calculate their interest feature. Initially, coordinates have been converted to decimal degrees with a 6 decimals precision (every 0.000001 decimal degree provides a 0.111 meters accuracy).

We would like to maintain this precision on every new feature that derives from geographical coordinates, like for example speed. Because some data mining algorithms (e.g. K-means clustering or FP-Growth) cannot process continuous datasets, the features need to be discretized. First we attend the set of geographical coordinates: we want to partition the whole length of the monitored bus route into smaller clusters. Each cluster should preferably cover up a road section of 25 - 50 meters, as we would like to approximate speed to this interval rather than to every couple of points. The size of the bins will have to vary according to the number of available example sets (use smaller bins - with higher precision, for a large number of examples or use wider bins for less data and consequently less accurate conclusions).The clustering operation can be accomplished with k-Means, where k will be a function of road length and the number of available example sets. Other features will be discretized by frequency: speed, altitude difference, time of the day. This operation takes as an input the continuous set of data and the number of bins to create, which is a function of the number of examples, the feature type and the expected result accuracy. Areas with high information density will contain narrower, more accurate bins, whilst sections with less data available will have wider bins. If there are large amounts of data available to us, we can afford a high number of bins; otherwise precision

will have to suffer. Attributes like day of the week and season of the year can be discretized by specification. For example, days can be divided into "workdays" and "weekends". A sample of such discretized features is depicted in Table II, where we stands for work day.

5.APPLYING ASSOCIATION RULES

We have selected and discretized the features that best relate to the speed variations on our track. Now we want to be able to draw some conclusions out of this information. We are going to use association rules. We have two suitable techniques to handle these situation as K-means association rule and FP-Growth Algorithm. The K-Means algorithm has Partitional clustering approach in which each cluster is associated with a centroid (center point) and each point is assigned to the cluster with the closest centroid.

The prerequisite for k-means is that one has to specify Number of clusters- K.

1.Select K points as the initial centroids.

2.Repeat

3.Form K clusters by assigning all points to the closest centroid.

4.Recomputed the centroid of the each cluster

5. until the centroids don't change.

The FP-Growth algorithm will help us identify frequent item sets.

Minimum support value for FP-Growth needs to be adjusted according to the number of available example sets, but also to the size of the bins generated through discretization and to the data reliability degree we are aiming for. We suggest using the following formula in order to estimate the minimum support:

$$minsup = \prod_{f=1}^{N} \frac{\alpha_f}{|bins_f|}$$

Where N is the number of selected features and |binsf | is the number of bins (partitions) created for feature f. Variable f depicts the degree at which examples are equally distributed within the bins of feature f. It takes values between 1 (the highest degree of equal distribution) and |binsf | (the lowest degree).If we strive to balance the number of examples among all time intervals and all road clusters, α_{time} and α_{road} should be closer to 1. However, there are other features to which this kind of approach would be nonsense. For example, it's highly impossible that we are going to get an equal distribution of all ranges of speeds over the same road section. Variable αspeed will probably be closer to $|bins_{speed}|$ in this case. After having generated frequent item sets, they are fed to the association rules operator, which takes the minimum rule confidence as an argument.

The result of this operation is a set of rules based on frequent item sets. Because our initial goal was to infer the causes of speed variations, we are only interested in rules that take speed as a conclusion. Provided that a sufficient volume of information exists, the presented system should be able to assess a set of reliable rules for predicting traffic behavior. Until now we were interested in associations that led to speed variations, but other valuable rules might also be inferred out of the knowledge base. For example we could determine the relationship between date & time values and passenger flow (or bus station interest).Potential benefits include:

1) Achieve a better coordination of the bus fleet: by knowing the correlations between the time of the day and the speed over each road section, buses running on the same route won't overlap their schedules anymore;

2) Offer a reliable bus schedule to both passengers and crew: the inferred rules could help put up a dynamic schedule on the company's website;

3) improve quality of service by determining high interest bus stations, aiming better coverage at critical hours;

4)balance fuel consumption and enable better duty scheduling and duty rostering ,uses smart-card information to determine the variability of public transit use, by the means of data mining techniques. Results describe the correlations between fare category and day of usage or boarding hours and frequently used bus stops.

A GPS data management system for GPS monitored buses, by deriving travel time patterns from historical data, and apply these patterns on real-time situations, with the corresponding adjustments. The authors focus on time values, whilst our system is based on time derived features, like speed, providing more knowledge than raw GPS data.

This is clearly not the intended effect, and there are currently a number of approaches, streams to deal with these problems under research such as *online optimization* which works on a zero knowledge assumption. The main drawback with this approach is that theoretical and practical performance doesn't match, *robust optimization*, and *stochastic optimization* this assumes complete knowledge on the likelihood of future events, which may be hard to obtain.

5.CONCLUSION

The theoretical assumptions related to the public transport system are evaluate in very different way when exposed to real time situation.

The Variations in traffic which includes all modes prevent us from assuming a fixed schedule over a certain bus route but we can trace out the common patterns that traffic follows and that these patterns can lead to accurate predictions of bus arrival and departure times. Ongoing work regards the combination of the above results with the activity theory to identify solutions for encouraging people to use public transportation.

REFERENCES

[1] Kedar R Joshi - RTI Application filed (12-Jan-2015) to PMPML Swargate Depot.
[2]Löbel "Optimal Vehicle Scheduling in Public Transit".
[3]Weider "Integration of Vehicle Scheduling in Public Transport".