Work Efficiency Prediction Analysis and Optimal Path Finding Algorithm

Vipul Ananda Bhangale
CSE dept., Dr. D.Y Patil Institute of technology, Pimpri-
Pune
Affiliated to SPPU
Pune, India

Rohit Radhakrishna Pillai
CSE dept., Dr. D.Y Patil Institute of technology, Pimpri-
Pune
Affiliated to SPPU
Pune, India

Bishal Anand
CSE dept. , Dr. D.Y Patil Institute of technology,Pimpri-
Pune
Affiliated to SPPU
Pune, India

Adarsh Kumar
CSE dept. ,Dr. D.Y Patil Institute of technology,Pimpri-
Pune
Affiliated to SPPU
Pune, India

Guide : Prof. Pradnya Tapkir – Kad
CSE dept., Dr. D.Y Patil Institute of technology, Pimpri-Pune
Affiliated to SPPU
Pune, India

Co-Guide : Prof. Vasudha Phaltankar
CSE dept., Dr. D.Y Patil Institute of technology, Pimpri-Pune
Affiliated to SPPU
Pune, India

Abstract **-- A correct prediction of a contractor's work can lead to many important things like keeping corruption in check, determining whether that particular is fit for the job or not. Frequently, it is brought that prediction is chaotic rather than random, which means it can be predicted by carefully analysing the history of respective contractor. Machine learning is an efficient way to represent such processes. It predicts a value close to the tangible value, thereby increasing the accuracy. The vital part of machine learning is the dataset used. The dataset should be as concrete as possible because a little change in the data can perpetuate massive changes in the outcome. It consists of variables like name of contactors, previous projects that he worked on, budget of that project, estimated lifespan of the project and actual life of the project.**

Due to the serious problem of bad road construction and maintenance, the real-time road situation and the possibility of road conditions in the next time period should be taken into accounts through the vehicle navigation system, in order to provide the optimal routing plans for vehicles in routing optimization. To solve the ignoring of real-time travel information and historical travel information in the existing navigation systems during routing optimization, this paper compares the existing navigation system to a prototype system which considers the conditions of the roadways as one of the factor in the recommendation and optimization system of the path finding and navigation process.

Keywords -- road resistance prediction; real-time navigation; historical travel information

1. Introduction

The main objective behind this system is to spread awareness about the road network among the citizens of the country. This system studies and empirically evaluates the roads predicament and the gives us an idea about what quality of work the construction unit has done and also provides us with a brief view of life of the roads and also the history of

the existing network of roadways. Apart from the analysis mentioned above, the system also tackles with a major issue of the existing web mapping and navigation services i.e. the services suggest a route that does not recognize the bad conditions of the roadways resulting in poor outcomes. The dataset should be as concrete as possible because a little change in the data can perpetuate massive changes in the outcome. It consists of variables like name of contactors, previous projects that he worked on, budget of that project, estimated lifespan of the project and actual life of the project. The path finding algorithm will not only consider the distance to calculate the best path but it will also consider the condition of the roads leading to one's destination, resulting in fast and efficient routes to travel.

The core content of map navigation software is path planning, and the essence of path planning is to find a suitable road according to the demands with the support of the algorithm. Due to the diversity of requirements, the application scenarios are also very rich. Such as emergency evacuation line design, logistics path planning and design, GPS-based navigation path planning, GIS-based road planning, robot optimal trajectory planning, UAV trajectory planning etc [1].

Existing path planning methods generally consider road length and the congestion on road, taking the shortest route from origin to destination as the navigation route. That is by using the classical algorithms to find the shortest path between two points, such as Dijkstra algorithm and Floyd algorithm. However, this kind of navigation method is existing problems: since the degree of damage of roads varies every couple of kilometres, the amount of time it consumes to pass this road is also different. Therefore, the shortest path is not necessarily the path with the least time consuming.

In recent years, domestic and foreign scholars have done a lot of researches on how to predict future time period's road conditions through various models and methods, such as the use of autoregressive moving average model, Kalman filtering model, neural network model, wavelet analysis, etc [2]. These methods and models are established based on statistics data. However, due to the urban travel system is a complex system that people participate in, it is time-varying and randomness, therefore making it difficult to use a mathematical model to predict the travel conditions in the future; Furthermore, if these statistics prediction methods are used widely in the path planning, it will cause more serious travel congestion [2]. The Real-time travel information collection is important for dynamic navigation with the increase of collected data, the database stores a lot of historical travel information which provides the foundation of the historical travel data analysis [3]. Therefore, we need to make predictions by using both real-time travel information and historical travel information.

In summary, to achieve the road travel path optimization, we need integrate road real-time travel information, road historical travel information and road quality information. By integrating road quality information, road real-time travel information and historical travel information, we design a comprehensive road resistance prediction model to calculate the road resistance. First, prediction model' s prediction effect is verified by experiment; then we use the prediction result as the road resistance, and design a path planning system to provide users with the travel route.

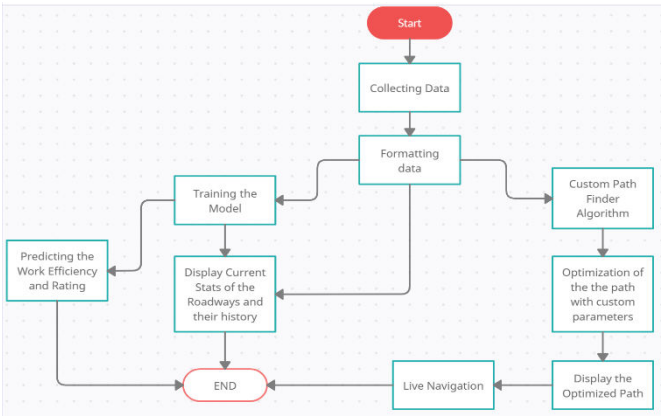
2. Data Collection (Surveys and Real-Time)

We collect most of the information about the history of the roads construction details from the government reserves and also need to run a survey area by area to generate the data about the current situation of the roads using a dedicated team initially but we have also provided a module which allows to users to update the situation of a particular road in real time. The system is smart enough to generate a good enough rating about the road's current condition based on the about information collected

The current navigation systems like Google maps and Safari do not collect this data at all. They are based upon the data of actual distance between the current location and the destination point of the user and also the congestion of traffic between the two points. Whereas, the prototype we propose collects the data of roads physical conditions like manholes, damage occur overtime and factor similar to that and helps the path finding algorithm accordingly.

The prototype maintains a database that stores this data that can be manipulated by verified users and he survey team dynamically. The existing system that are currently under use by most of the public do not have such provisions.

3. Project Efficiency Prediction Model



Prediction of state of the road is very challenging and complicated process because it depends on multiple factors and is time varying. The factors being the weather conditions, the movement of heavy vehicles etc. Although the life of road is pre-decided at the time of estimation, it erodes way before the lifetime speculated earlier. We will be using Random Forest algorithm in order to combine these factors together to train the prediction model.

First, the random forest algorithm is used to order feature importance and reduce dimensions. Second, the selected features are used with the random forest algorithm and the F-measure values are calculated for each decision tree

as weights to build the prediction model for project analysis [8].

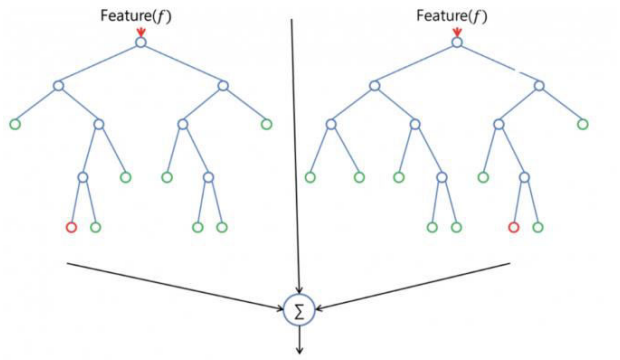
The system design is such that, the user visiting the particular road is given an option to rate the road on a scale of 1 to 5. The rating is further used to calculate the rating of the concerned contractor at real-time. The overall ratings of the previous works and the real-time data feed is passed on to the ML module of the system for analysis and result prediction.

4. Random Forest Algorithm

Random forest is a flexible, easy to use machine learning algorithm that produces, even without hyper-parameter tuning, a great result most of the time. It is also one of the most used algorithms, because of its simplicity and diversity. It can be used for both classification and regression tasks.

How it works. Random forest is a supervised learning algorithm. The "forest" it builds, is an ensemble of decision trees, usually trained with the "bagging" method. The general idea of the bagging method is that a combination of learning models increases the overall result. Simply, random forest builds multiple decision trees and merges them together to get a more accurate and stable prediction.

One big advantage of random forest is that it can be used for both classification and regression problems, which form the majority of current machine learning systems. Let's look at random forest in classification, since classification is sometimes considered the building block of machine learning [9]. Below you can see how a random forest would look like with two trees:



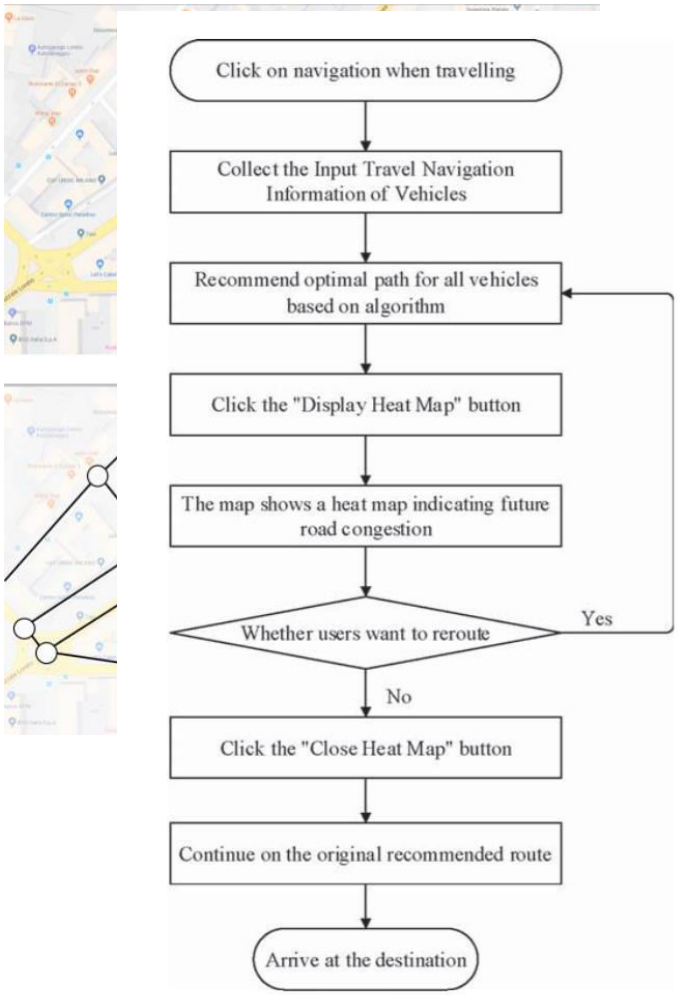
5. Path Finding Algorithm

Under the guidance of application scenarios, the corresponding algorithms are also constantly optimized. The most classic algorithm was proposed by Dutch scientist Edsger Wybe Dijkstra in 1959, and many improved algorithms were completed on this basis subsequently. For example, A* heuristic search algorithm, the literature [4] systematically explained the whole process from the definition of A* algorithm to the proof of algorithm optimality. Rajotia S. [5] simulated the bidirectional path with heuristic algorithm. The results showed that the proposed method can reduce the vehicle blocking time and improve the throughput capability of the system. Aloquili O. [6] used the Dijkstra algorithm and the Kruskal algorithm to display the real position of the vehicle through real-world geographic information technology, and made decisions based on real-time information. Geisberger Ret al. [7] proposed a road network shrink hierarchy suitable for mobile devices, improving the performance of many-to-many routing, cross-node routing and other scenarios. In recent years, various group intelligent algorithms, such as ant colony algorithm and genetic algorithm have also brought new solutions to path planning. Google Maps (one of the most used navigation tool used) is based on a very simple but incredibly effective algorithm: the Dijkstra algorithm.

What does Dijkstra's algorithm do? Given a weighted graph, a starting point and an endpoint within the graph itself, the algorithm finds the "minimum path" that connects the two points, that is the sequence of arcs that minimizes the sum of the weights and therefore, in the case of Maps, minimizes the estimated travel time.

What does Prototype's Custom Path Finding algorithm do? The algorithm assigns a weight that corresponds to the condition of the road in real time to the arcs that join the two points i.e. starting location and the destination point. These weights are calculated through the data we pre-process in the database that supports the algorithm in the backend.

How a map becomes a graph. The streets are arches (the black lines), while the intersections are nodes (the white circles):



The arches are not all the same**.** When we have to choose between two possible road we take into account the one that gets us to destination first. The factors that come into play are many: the maximum speed, the size of the road, the presence of traffic lights, traffic and so on. We can summarize them all in a single value, the (estimated) travel time of that section. This information is integrated into our graph by weights, that are the values attributed to each arc. Consequently, if a "4" is written on the arc that connects the node X and the node Y, this indicates that to go from the intersection X to the intersection Y we estimate that it takes 4 minutes.

In order to facilitate the data processing and reduce pressure on the server, the real-time travel information is not collected continuously, but collected at every certain time interval. If the interval is set too large, the real-time information will change significantly within this interval, and such information is not real-time information anymore; if the interval is set too small, the real-time information will update too frequently, and it will not only increase the burden of the server, but also make the data fluctuant.

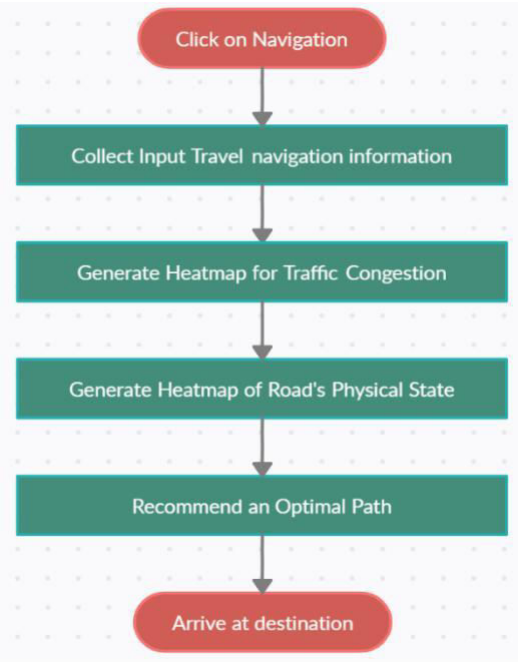
Hence, using these customized weights we have improved the simple path finding algorithm's input such that comprehensive cost of travelling from one node to another node changes according to the conditions of the roads. The next section of the article mainly introduces the travel planning model, including the implementation of the ideas and the improvement of the algorithm.

6. Travel Planning Model

Existing systems Planning Model: These systems generate a Heat map that represents the congestion on the roads. When describing the running state of traffic roads, there are three important parameters, namely speed, density, and flow. A lot of research has been done on traffic flow models at home and abroad.

The simple flow chart of how the congestion information is processed and implemented to obtain optimal routes is displayed as follows:

Proposed System Planning model: Compared with the real-time traffic function of the existing map software, this app is unique in its ability to analyse the ongoing road conditions based on the real-time travel plans data from the travellers and based on historical data statistics, the data availability is stronger. In addition to the heat map generation that represents the congestion of traffic vehicles, the system generates another heat for physical state of the road and allows user to optimize the route according to time takes in order to cross a nicely paved road vs a damaged road.



Drivers start from the current time T, and according to the conditions set, the navigation is enabled at the departure. The optimal path planning is carried out by the improved Dijkstra algorithm, and the real-time routes of all navigation vehicles are calculated and analysed based on the cloud computing platforms. Through the congestion indicator of the heat map as well as the physical state indicator of the heat map, drivers decide whether to change the travel routes according to their own needs, thereby avoiding congestion.

The Travel Speed Influenced by Road Quality: The quality of road includes the road level and the road speed restriction information. The urban road is mainly divided into four levels: the fast road, the trunk road, the secondary trunk road and the slip road. The classification criteria for each are as follows:

- The fast road: The fast road whose width is greater than 40 meters, having a median strip in the middle, with four or more lanes, is for cars to travel fast. Its designed speed is 60km/h-80km/h. z
- The trunk road: The trunk road whose width is between 30 m and 40 m, is used for connecting the city districts, the main function of it is transportation. Its designed speed is 40km/h-60km/h.
- The secondary trunk road: The secondary trunk road whose width is between 20m-24m, and its main function is to undertake distribution between trunk road and the districts and has a service function. Its designed speed is 40km / h.
- The slip road: The slip road, whose width is between 14m and 18m is used for connecting road between the secondary road and the neighborhood road, whose main function is service function. Its designed speed is 30km / h.

For road without clear speed restriction information, we take the maximum design speed as its free-flow travel speed; for road with clear speed restriction information, we take the restricted speed as the travel speed. We mark it as Vi1.

$$V_{i1} = \begin{cases} \text{max design speed} & \text{without speed restriction;} \\ \text{restricted speed} & \text{with clear speed restriction;} \end{cases}$$

In order to be stored and used easily, the fast road, the trunk road, the secondary trunk road and the slip road will be stored as 1,2,3,4 grades in the database.

It can truly make the road ahead transparent in the future time, thus improving the smoothness of roads, alleviating the current situation of urban traffic congestion and physical conditions, and helping travelers to enhance their travel experience.

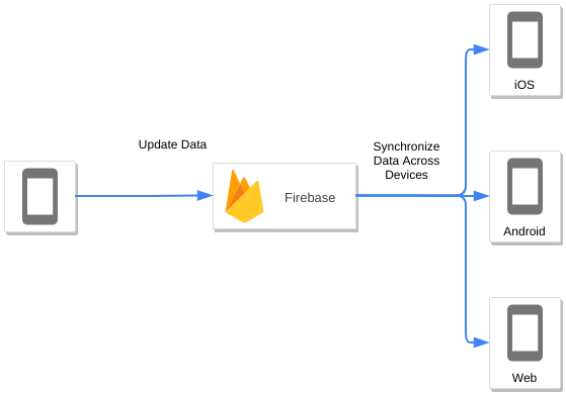
7. Data Operations Management

DataOps (data operations) is an Agile approach to designing, implementing and maintaining a distributed data architecture that will support a wide range of open-source tools and frameworks in production.

How DataOps works. The goal of DataOps is to combine DevOps and Agile methodologies to manage data in alignment with business goals. For example, if the goal is to raise lead conversion rate, DataOps would position data to make recommendations for marketing products better, thus converting more leads. Agile processes are used for data governance and analytics development while DevOps processes are used for optimizing code, product builds and delivery [10]. We will be combining the concepts of DataOps with the database for android called as Firebase. The web application has become more and more reliant upon large amount of database and unorganized data. It is difficult for Relational Database Management System (RDBMS) to handle the unstructured data. Firebase is a relatively new technology for handling large amount of unstructured data. It is very fast as compared to RDBMS. Firebase came into existence for Android apps which uses JSON for storing data. The other servers use a table (rows and columns) format for storing data. Firebase is NoSQL based. There are very few cloud-based server available which are similar to firebase [11].

8. Database Architecture

Firebase Realtime Database is a Cloud hosted database, i.e. it runs on a cloud and access to the user is provided as a service. It stores data in JSON (JavaScript Object Notation) format, a format to store or transport data. All the users connected to it can get access to the data at Real Time. Hence, with real time connection, our users will be able to provide our system with data needed to keep the states of the roads up-to-date.



Realtime Database (RTDB) stores the data as JSON tree. It offers many client SDKs from iOS, Android, and Web. The database of the system consist of three Schemas – one that holds the information about the user which consist of fields like email id, user name, authentication details, gender etc, another one stores the data about the roads, their addresses, the ratings and the last one consist of the data about the contractors and the rating calculated from the second table.

The machine learning algorithm takes input from the second table i.e. the fields roads costs, expected life, current state etc as the dependent variables and hence decides whether the particular contractor must be given the upcoming project or not.

9. Conclusion

The goal of this paper is to provide users with the route with the least time consuming. Integrating the road quality information, the road real-time travel information, the road historical travel information, we designed a road resistance prediction model which takes the road travel time as the road resistance. The experimental result shows that the prediction model has good predictive effect for the users to select the optimal path in the rush hour. The prediction model can provide the reference for mobile navigation devices and car navigation devices in path planning. This new idea increases the probability of traveling smoothly, and provides accurate decision-making basis for traffic control.

At present, due to data barriers, barely accurate data can be collected for research. Hence, we have used dummy data in a small local area. However, in the future, the transparency of traffic data is the direction of the national big data policy.

At the same time, this paper will adopt a more comprehensive model to deal with the more complicated actual situations in the subsequent research, so as to achieve more accurate display of the road conditions ahead. It not only provides a new method to solve the problem of urban congestion and travel time optimization, provides a good research foundation and direction for vehicle travelling, but also contributes to the construction of intelligent transportation and intelligent city based on big data.

10. Acknowledgement

We would like to acknowledge the contributions of the following people without whose help and guidance this project would not have been completed. We acknowledge the counsel and support of our project guide coordinator, Mrs. Pradnya Tapkir - Kad, Professor, CSE Department, with respect and gratitude, whose expertise, guidance, support, encouragement, and enthusiasm has made this project possible. Their feedback vastly improved the quality of this project and provided an enthralling experience. We are indeed proud and fortunate to be supported by her. I am also thankful to Prof. (Dr.) S.V. Chobe, H.O.D of Computer Science for his constant encouragement, valuable suggestions and moral support and blessings. Although it is not possible to name individually, I shall ever remain indebted to the faculty members

11. References

[1] WANG Jian, HU Xiao-wei, TONG Jin-jin, HE Lei. Route planning of regional emergency evacuation based on lane modeling [J]. Journal of Traffic and Transportation Engineering, 2010, 10(02): 82-87

[2] Lu Feng, Duan Yingying, Zheng Nianbo, “A practical route guidance approach based on historical and real-time traffic effects”. State Key Laboratory of Resources and Environmental Information System, IGSNRR, Chinese Academy of Sciences.

[3] Qihui Qin, Meiling Feng, Junqing Sun and Bin Sun, "Prediction of road resistance based on historical/real-time information and road quality," 2015 12th International Conference on Fuzzy Systems

and Knowledge Discovery (FSKD), 2015, pp. 1073-1077, doi: 10.1109/FSKD.2015.7382091

[4] Dechter R, Pearl J. Generalized best-first search strategies and the optimality of A* [J]. Journal of the ACM , 1985, 32(3): 505-536.

[5] Rajotia S, Shanker K, Batra J L. A semi-dynamic time window constrained routeing strategy in an AGV system [J]. International Journal of Production Research, 1998, 36(1): 35-50

[6] Aloquili O, Elbanna A, Al-Azizi A. Automatic vehicle location tracking system based on GIS environment [J]. IET software, 2009, 3(4): 255-263.

[7] Geisberger R Sanders P, Schultes D, et al. Exact routing in large road networks using contraction hierarchies [J]. Transportation Science, 2012, 46(3): 388-404.

[8] Xiang Gao, Junhao Wen, Cheng Zhang, "An Improved Random Forest Algorithm for Predicting Employee Turnover", *Mathematical Problems in Engineering*, vol. 2019, ArticleID 4140707, 12 pages, 2019. https://doi.org/10.1155/2019/4140707

[9] https://builtin.com/data-science/random-forest-algorithm

[10] https://searchdatamanagement.techtarget.com/definition/DataOps

[11] Chunnu Khawas, Sikkim University, International Journal of Computer Applications 179(46):49-53 DOI: 10.5120/ijca2018917200