Introduction

Traffic congestion is due to increased need where the availability is less. So vehicle traffic congestion has become one of the key issues in today’s society. It is being experienced everyday by millions of people in day-to-day life. Due to traffic congestions many people face problems such as increased travel time, mental stress, accidents, wastage of resources and many others. Though there were several physical measures that were taken to reduce traffic like proper road infrastructure, urban planning and design, and other traffic management practices this did not completely solve the issue. This eventually led for a need of an effective and efficient software system that will reduce the traffic chaos and will eventually lead for reduced travel time. In this paper we propose a system that deals with an efficient prediction method that will reduce traffic congestion and allows us for optimal usage of resources.

Series of clustering and prediction techniques is applied on the data set and has been used to predict the traffic over a road point and by finding the similarity value we can quantify the influence of that road points over others. The major benefit of this system is that it reduces the traffic chaos and will eventually lead for reduced travel time. In this paper we propose a system that deals with an efficient prediction method that will reduce traffic congestion and allows us for optimal usage of resources.

Related works:

A. Data Preprocessing:
Step 1: Collect the traffic detail for all road points.
Step 2: If the collected traffic values have missing values then do step 3 else do step 9.
Step 3: Find those missing values.
Step 4: Now fill those missing values.
Step 5: Compute the relative value that has to be filled.
Step 6: Fill all the missing values of the raw dataset.
Step 7: Apply the clustering technique on the data set.
Step 8: Predict the traffic congestion using the clustering technique.

Proposed System Architecture

Fig. 1. Architecture of the Proposed System

In this paper [3], he used the binary neural network for traffic prediction which gives similar accuracy like similarity value calculation. Even though the system reduces the accuracy, it is more reliable in all the traffic condition. J.Hall and P.Mars[4] calculated the traffic values by using the time series method which gave a better result even though they found some performance issues. In [5][6][7], the mining technique used for prediction and clustering is done with the help of historical dataset. In this paper[8] they compared the multivariate and univariate approaches to forecast the traffic condition. In [9][10], They have used the historical data as well as the current data to predict the traffic level in accurate level.

Proposed System

A. Data Preprocessing:
Step 1: Collect the traffic detail for all road points.
Step 2: If the collected traffic values have missing values then do step 3 else do step 9.
Step 3: Find those missing values.
Step 4: Now fill those missing values.
Step 5: Compute the relative value that has to be filled.
Step 6: Fill all the missing values of the raw dataset.

They used corridor travel time method which reduced the reliability, later the author suggested using as link based method.
Step 7: Remove the noisy data
Step 8: From the collected values extract values that are relevant to the road points based on area code.
Step 9: Based on the area code store them in separate data mart.

The first phase of this paper discusses on data preprocessing algorithms used to clean raw log data. Data preprocessing helps to extract certain data from raw data which in term transformed into other necessary form. By cleaning the data, we can create the database according to our application which includes the information only about traffic at each road points.

Gather the traffic data of different road points that has been accumulated over years. Now these data may contain data that are not relevant to that road point, it may also contain missing values or noisy data. These various issues have to be addressed so we have to apply appropriate preprocessing algorithm to clean and extract data.

Then our raw dataset doesn’t have any missing values, but it still have irrelevant data and noisy data. In order to remove these data we can extract only the relevant data by using the area code, and form them as separate data marts.

B. Prediction:
Step 1: Get valid road point and day from user.
Step 2: Load dataset of given road point on given day.
Step 3: For other road points load the dataset of the particular day.
Step 4: Compute the dissimilarity of that given roadpoint with other roadpoints.
Step 5: Find the dissimilar roadpoints.
Step 6: Eliminate the dissimilar roadpoints.
Step 7: Collect the similar roadpoint data.
Step 8: Based on these variables compute the traffic value.

The preprocessed data has been used to predict the traffic. Prediction module gets the data finds the dissimilar values, eliminates the dissimilar roadpoints, gather the similar roadpoints value and find the traffic value of the given roadpoint these value is verified and send to clustering module for further clustering.

Prediction algorithms analyze the historical data which is quite similar to present condition and mark it as y. Then, “given a n records (each of which is described by K real-valued attributes) and a dissimilarity measure dm, [nearest neighbor searching] finds the y road points closest to road point rp”.

C. Clustering:
Step 1: Get the computed traffic value.
Step 2: Check the traffic value of neighbor road points.
Step 3: Find the maximum fs, where s is the shortest distance from p to q.
Step 4: Compute Fq’ where Fq’ is Fq(traffic condition of given road point)+ sum of difference between the traffic in and traffic out.
Step 5: Similarly compute Fp’ for other road points.
Step 6: Find the distance between the given road point and other road points.
Step 7: Finally compute similarity value.
Step 8: Repeat the above steps to compute similarity with other road points.

The traffic at particular road point not only affects the traffic over that point it may also affects other points and also similar case may be predicted.

We define $S_{pq}$, the similarity value is calculated by identifying the probability of q is influenced with the assumed p based on the input data where $f_p = \text{mod}(p + \sum N_i \text{in} \times \sum N_o \text{out} \times f_n)$, and $f_{maxpq}$ is the maximum $f_s$, with s is the road point on the shortest distance from p to q. If the similarity value of particular road point p is high, then q may be highly influenced on p. Every road point will have the high influence on itself.

Consider if $p=q$, then the similarity value is assumed to be 1. And if $p \neq q$, then the similarity value will be less than 1. Here the road point affects the other road point only if it has high influence over it and $S_{pq}$ is directly proportional to the $f_q$. The road point which has complex data may highly influence on the neighbor point. The less complex data may not have much influence on it rather it is influenced in some other road point. The complexity of traffic data is relatively proportional to the volume of the chosen data. Only due to the high volume of data, the traffic congestion occurs. The self influenced point is considered as a high traffic area. Clustering the road points based on the similarity that we analyzed over each road point. Assume if the road point q is less influenced over p then it has low similarity value where the point p is far away from q.

Result:
In this proposed system, we found a way to reduce the traffic congestion that determines the traffic condition and affected road points. The above algorithm helps to reduce the travel time of the user. Beyond accessing the statistical data, the user would gain the opportunities to plan their travel. Mainly designed to avoid the accidents and reduce the travel time. The user can plan the overall journey. This is hoped to make some significant change in the safety road travel.
Future Work:
In future, the performance of the algorithms is tested with large amount of traffic data. We will design to accept more inputs from user to predict the more accurate result that can able to withstand in any traffic condition. The sub-road points displayed in the output screen will provide a path to travel which is to avoid deviation from the actual way. Further this system can be extended to send updates about the traffic condition to the user’s mobile and can be improved to compute travel time, allowed speed, alternative travel paths.

Reference: