Decision Tree Analysis to Predict Traffic Congestion in Transport Routing

J. Patricia Annie Jebamalar, Dr. Sujni Paul and Dr. D. Ponmary Pushpalatha

Abstract--- Geospatial information system (GIS) provides a good support for users who are in search of optimal route, as searching the optimal path is one of the advanced analysis functions of GIS. The optimal route finding in Intelligent Transport Routing (ITS) is a challenging task. Road users want reliable door-to-door trips that are free of stress. Much can be done to reduce the worst traffic congestion This research aims to enhance transport routing mechanism to be most efficient and strong to face the most important routing problem, traffic congestion. This research aims to enhance transport routing mechanism to be most efficient and strong to face the most important routing problem, traffic congestion. To solve the problem of traffic congestion, the existing road routes are analyzed by classifying them according to the congestion level. A decision tree is derived which will help the users to predict the level of traffic congestion.

Keywords--- Transport Routing, Classification, Decision Trees and Traffic Congestion

I. INTRODUCTION

Transport Routing consists of two fundamental steps moving vehicles from one place to another (from source to destination) and determining which path to follow (specifying a route). Traffic congestion could be a condition on road networks that happens as use will increase, and is characterized by slower speeds, longer trip times, and increased transport queuing. The most common example is the physical use of roads by vehicles. When traffic demand is high that the interaction between vehicles slows the speed of the traffic stream, traffic congestion is caused. Traffic congestion is often the outcome of successful economic development, employment, housing and cultural, policies that make people want to live and work relatively close to each other. Congestion prevents us from moving freely. It is difficult to see how congestion can or should be eradicated. Congestion control refers to the set of actions taken by the transporter to minimize the travel time.(Ouri Wolfson, Bo Xu)

Data mining is defined as the process of discovering patterns in data. The overall goal of the data mining process is to extract information or patterns from a data set and transform it into an understandable structure for further use. The patterns discovered must be meaningful in that they lead to some advantage. This paper suggests a congestion control policy (Soukaena 2011) by deriving decision trees from the road database which can be analyzed to predict the congestion level.

II. RELATED WORK

Managing Urban Traffic Congestion is a Report By European Conference of Ministers of Transport (ECMT) in 2007 aimed at those in charge of preparing congestion management policies and also those responsible for improving congestion management operations. It aims to provide a better understanding of the phenomenon of congestion and provide guidance in relating this understanding to local circumstances. (ECMT 2007) Adaptive Routing for Road Traffic (IEEE 2000) is about the

J. Patricia Annie Jebamalar, M.Sc., M.Phil.,Research Scholar, Karunya University, Coimbatore, Tamilnadu, India. E-mail:patricia.benlaz@gmail.com
Dr. Sujni Paul, M.C.A., Ph.D., Professor, The Oxford College of Engineering, Bommanahalli, Bangalore, India.E-mail: sujnidejus@gmail.com
Dr. D. Ponmary Pushpalatha, M.C.A., M.Phil., Ph. D.,Professor, Karunya University, Coimbatore, Tamilnadu, India. E-mail:ponmary@karunya.edu

software has demonstrated that adaptive computerized route generation could help drivers cope with the growing congestion (John Fawcett and Peter Robinson 2000). Applying Data Mining Techniques for Traffic Incident Analysis (Journal of The Institution of Engineers, 2004) discusses computer-based simulation and visualization tools to evaluate new algorithms for incident detection and strategies for incident management. Applications of such tools change users to gauge solutions quicker than real-time. However, it looks that there is still scope to better investigate things like traffic incidents and study their impacts, additional completely by using techniques like data mining. (Der-Horng Lee, Shin-Ting Jeng and P. Chandrasekar 2004) Towards Fine-Grained Urban Traffic Knowledge Extraction Using Mobile Sensing (ACM 2012) introduces the vision for mining fine-grained urban traffic knowledge from mobile sensing, especially GPS location traces (Xuegang (Jeff) Ban, Marco Gruteser 2012). Urban Traffic Intersection Incident Prediction Using AI Algorithm presents a new method using data mining to identify automatically freeway incidents.(Yaguang Kong, Huakui Chen) Using Data Mining to Forecast Uncertain Demands in Stochastic Vehicle Routing Problem (ISEP 2005) says, in the field of vehicle routing, relatively limited attention has been paid to the topics such as stochastic demand. With the data warehouse infrastructure and data mining tools it is possible to predict such variable constraints. This could modify a lot of efficient route construction and as consequence the amount of non accomplished routes could be reduced. (Hrvoje Marković, Ivana Ćavar, Tonči Carić 2005)

III. THE DESIGN OF PROPOSED POLICY

The proposed system aims to build an efficient congestion control policy. The design of the proposed policy aims to predict the level of congestion in road network. Initially, the traffic data is preprocessed to get the reliable data. Then this proposal finds out the level of traffic congestion on transport routing network by classifying the database and deriving a decision tree. This classification is performed by using decision tree data mining technique. The derived decision tree helps us to predict the level of traffic congestion. The following sections will explain the proposed system in detail.

A. Preprocessing Stage

This stage presents the initial step in the proposed policy which will consider the reliable database. The following procedure presents the overall process of the preprocessing stage:

a. Preprocessing Steps

Input: Get all reliable road links and the related data such as distance, congestion level, hour of the data and travel time.

Output: Proposed Congestion Control Policy database.

Step 1: Construct the structure of the proposed database.

Step 2: For each route insert the attributes values into the table.

Step 3: End.

B. Decision Tree Mining Stage

Decision Trees are considered to be one of the most popular approaches for representing classifiers. (Han, J., Kamber 2006). A decision tree is a classifier expressed as a recursive partition of the instance space. In a decision tree, each internal node splits the instance space into two or more sub-spaces according to a certain discrete function of the input attributes values. In the simplest and most frequent case, every test considers a single attribute, such that the instance space is partitioned per the attribute's value. Each leaf is assigned to one class representing the most appropriate target value. Decision tree learning is a method commonly used in data mining. The goal is to form a model that predicts the value of a target variable based on several input variables. In this research work the target variable is congestion.

a. J48 Decision Trees

C4.5 is an algorithm used to generate decision trees from a set of training data. This algorithm has been implemented in Java and has been included in Weka, an open source data mining tool, as J48. A decision tree is a predictive machine-learning model that decides the target value (dependent variable) of a new sample based on various attribute values of the available data. The internal nodes of a decision tree denote the various attributes, the branches between the nodes shows the possible values that these attributes will have within the observed samples, whereas the terminal nodes shows the final value (classification) of the variable.

The attribute that is to be predicted is known as the dependent variable, because its value depends upon, or is determined by, the values of all the other attributes. The other attributes, that facilitate in predicting the value of the variable, are called the independent variables within the dataset.

The J48 Decision tree classifier follows the following simple algorithm.

Input: Proposed Congestion Control Policy database.

Output: Decision Tree to Propose Congestion Control Policy.

Step 1: Create a decision tree based on the attribute values of the available training data.

Step 2: Whenever it encounters a set of items (training set) it identifies the attribute that discriminates the various instances by choosing the instance that has highest information gain.

Step 3: If there is any value for which there is no ambiguity, that is, for which the data instances falling within its category have the same value for the target variable, then terminate that branch and assign to it the target value that is obtained.

Step 4: Otherwise, look for another attribute that gives us the highest information gain.

Step 5: Continue until a clear decision of what combination of attributes gives a particular target value is got, or run out of attributes.

C. Proposed Congestion Control Policy

Preprocessed traffic data is given as input to the J48 classifier of the Weka tool and the decision tree is derived. The leaf nodes of the decision tree denote the level of congestion. By analyzing the internal nodes carefully the level of congestion is decided by observing the dependent variables such as source, destination and hour of the day. This helps to predict the level of congestion. The derived decision tree can be converted into a graph structure which can be used to decide the optimal path between a particular source and destination based on the hour of the day.

IV. DISCUSSION AND EXPERIMENTAL RESULTS

The problem of transport network congestion control remains a critical issue and a high priority, especially in given growing size of the transport networks. Congestion is a real threat to the growth of existing transport networks, and of the future development of transport networks. It is a problem that cannot be ignored. From that the research concentrates on enhancing the policy to deal with congestion. The implementation of the proposed system will begin with taking samples of transport network and then by the proposed congestion control policy, which tends to track and record all details related to the path from source to destination,

The following table presents the proposed database which will be introduced to decision tree data mining technique,

- Attribute 1 : source
- Attribute 2 : destination
- Attribute 3 : hour
- Attribute 4 : congestion

The experimental analysis begins by loading the data into WEKA, as seen in Fig.1. Next, select the "Classify" tab and click the "Choose" button to select the J48

classifier. Now click on "Start" to generate the decision tree.

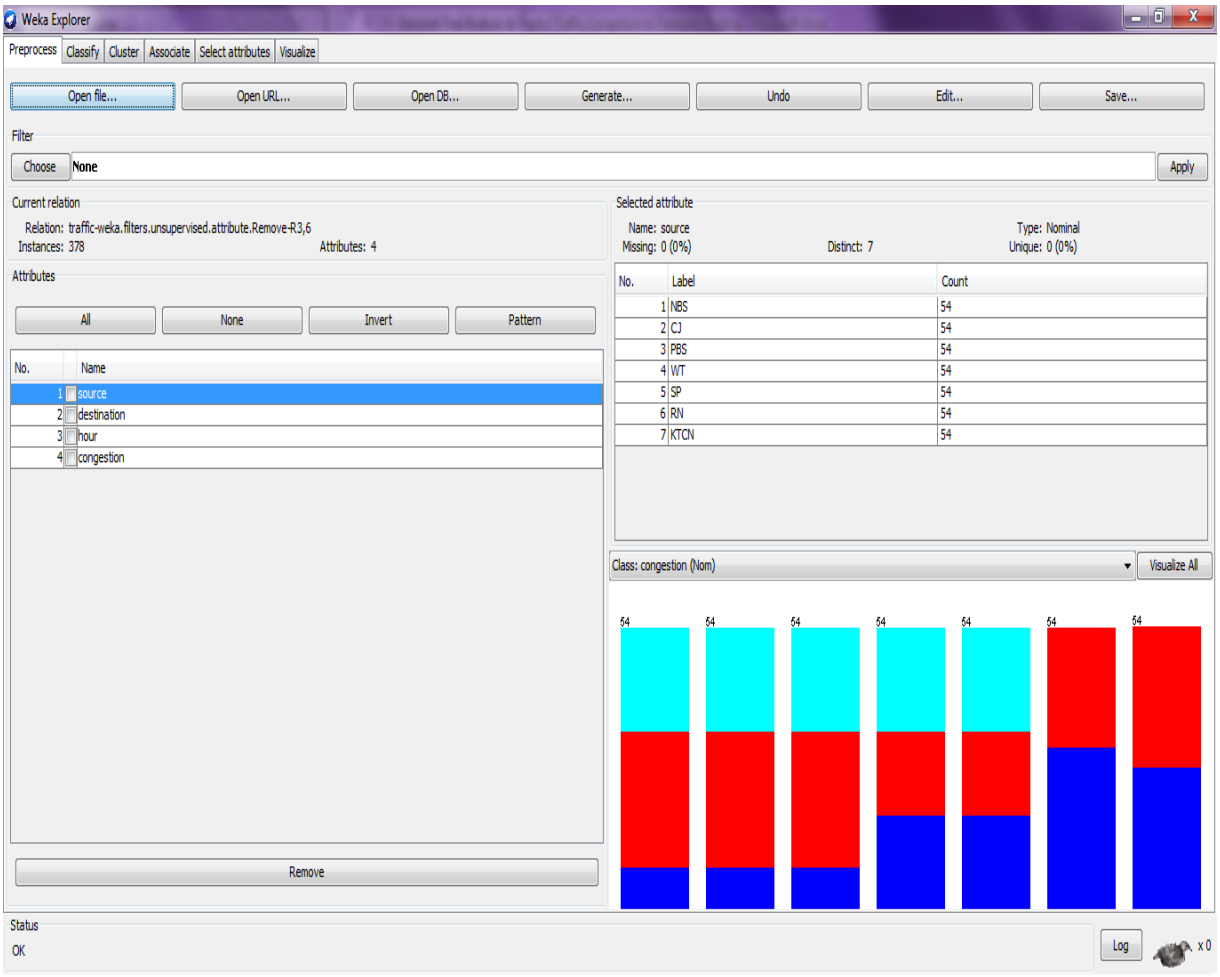

Fig. 1: Loading the dataset

To view a graphical rendition of the classification tree right click the last result set and select "Visualize tree" from the pop-up menu. The tree for this experiment is depicted in Figure (2). Based on the source, destination and hour of the day, the level of congestion is measured as low, medium and high.

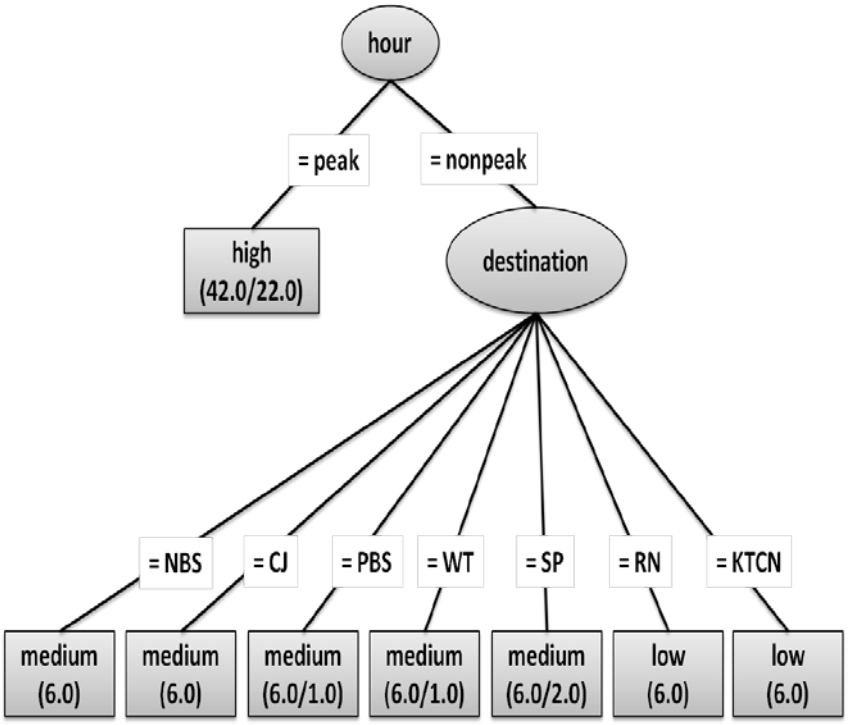

Fig. 2: The Decision Tree Showing the Level of Congestion

V. CONCLUSION

This paper has presented decision tree analysis which aims to enhance transport routing mechanism to be most efficient and strong to face the most important routing problem, traffic congestion. The experimental result shows that the derived decision tree helps to predict the level of congestion. Based on the level of congestion, the transport routes can be planned in a better way. In future, the derived decision tree can be analyzed further by converting the decision tree into a parameter while deriving the optimal path.

REFERENCE

[1] Soukaena Hassan Hasheem, An Association Rules Analysis to enhance solving the Congestion Problem, University of Technology, Iraq, Journal of Emerging Trends in Computing and Information Sciences, Volume 2 No.7, JULY 2011

[2] Ouri Wolfson, Bo Xu, Spatio-temporal Databases in Urban Transportation, University of Illinois, Chicago, Bulletin of the IEEE Computer Society Technical Committee on Data Engineering

[3] Neal Lathia, Jon Froehlich, Licia Capra, Mining Public Transport Usage For Personalised Intelligent Transport Systems

[4] A.John Sanjeev Kumar, J.Arunadevi, V.Mohan, Intelligent Transport Route Planning Using Genetic Algorithms in Path Computation Algorithms, Thiagarajar College of Engineering Madurai, TamilNadu, India, European Journal of Scientific Research, 2009

[5] S. Behin Sam, Amna Omar, A. Kannan and P. Viekanathan, The Application of Data Mining Technologies for Routing in Communication Networks, Asian Journal of Information Technology 2006

[6] Han, J., and Kamber, M. “Data Mining: Concepts and Techniques”. Morgan Kaufmann, 2006.

[7] Managing Urban Traffic Congestion, a report by European Conference of Ministers of Transport (ECMT) in 2007

[8] John Fawcett and Peter Robinson, Adaptive Routing for Road Traffic, University of Cambridge, IEEE 2000

[9] Der-Horng Lee, Shin-Ting Jeng and P. Chandrasekar, Applying Data Mining Techniques for Traffic incident analysis, Journal of The Institution of Engineers, Singapore, Vol. 44 Issue 2 2004

[10] Xuegang (Jeff) Ban, Marco Gruteser, Towards Fine-Grained Urban Traffic Knowledge ExtractionUsing Mobile Sensing, ACM 2012

[11] Yaguang Kong, Huakui Chen, Urban Traffic Intersection Incident Prediction Using AI Algorithm, Institute of Information and Control, Hangzhou Dianzi University,310037,China

[12] Hrvoje Marković, Ivana Ćavar, Tonči Carić, using Data Mining to Forecast Uncertain Demands in Stochastic Vehicle Routing Problem, Faculty of Electrical Engineering and Computing, Unska 3, 10 000 Zagreb, Croatia, Faculty of Transport and Traffic Engineering, Vukelićeva 4, 10 000 Zagreb, Croatia, ISEP 2005