Using Data Mining Techniques on Fleet Management System

Chang-Yi Chen, Tien-Yin Chou, Ching-Yun Mu, Bing-Jean Lee,
Magesh Chandramouli, Hsien Chao
Geographic Information Systems Research Center, Feng Chia University
100 Wenhwa Rd., Taichung, Taiwan, R.O.C.
Tel: +886-4-24516669 ext. 59 Fax: +886-4-24519278
E-mail: {Daniel, jimmy, jackie, bjlee, magesh}@gis.fcu.edu.tw, k4228@tccg.gov.tw

Abstract

Traditional transportation vendors not only strive to deliver cargo securely and accurately to customers on time, but also consider reducing the operation cost and flexibly dispatching vehicles as well as staff. Thus, how to raise the competitive advantage is the big challenge for the transportation industry. Because the online vehicle tracking system can accumulate enormous records of vehicle operations management of the information is beyond the range which could be handled by human ability. The paper will discuss how to explore the useful data of vehicle behaviors from enormous records of vehicle operations by data mining. The data of vehicle behaviors can help us understand the status of every vehicle or driver, such as being out on duty, driving against traffic regulations, and deviating from routes. It can also alert to abnormal conditions, releasing the burden of fleet management.

Index Terms - Spatial Data Mining, Geographic Information Systems,
Global Positioning System, Fleet Management.

Introduction

The incessant economic and industrial activities around the globe and the splurge of exports and imports continue to impose greater demands on the shipping and cargo industry. The prompt and secure delivery of consignments is no longer the sole objective of cargo deliverers. Shipping companies are beginning to focus on the reduction of operation costs and flexible dispatch of vehicles and staff. In spite of the increase in the number of consignments handled by shipping companies, the profits have been plunging down recently. Hence, besides offering competent services to their clientele, transportation industries also need to manage their operations profitably. Information technology holds the key to this crisis and hence transportation vendors have turned their attention to e-Commerce and m-Commerce. The recent developments in mobile communication and the advances in GPS and GIS can be exploited to build a Fleet Management System for online tracking and dispatch of vehicles.

Our fleet management system consists of a car-kit with a wireless communication set-up that is installed in every vehicle, and a centrally located monitoring platform or control station. The vehicles are equipped with GPS and wireless communication devices that automatically transmit the vehicle details, such as coordinates and speed to the display in the central monitoring station. Hence, a driver simply needs to press a button while loading or unloading cargo. The administrator or supervisor monitors the present status of the vehicle and the cargo through the map display at the central monitoring platform. The system automatically alerts the supervisor about instances of over-speeding, illegal parking, or other emergencies, and the supervisor can deal with the situation in the earliest possible time. The system provides GSM Short Messaging Service, using which the supervisor can send instructions to the drivers, make emergency dispatches, and circulate other important communications. In addition, the supervisor can utilize the automatic dialing facility to communicate on a one-to-one basis with a specific driver. While shipping essential or dangerous commodities, the supervisor can also keep an eye on a particular vehicle via individual displays. The system also automatically generates a report that helps in the documentation of vehicle status over a period of time.

Theoretically, for accurately tracking the status of all the vehicles, a detailed record of speed, direction, and location for each moment is required. Assuming that one record is generated per second, this amounts to 28,800 records for each vehicle for 8 hours a day. A large fleet that includes more than 500 vehicles can generate 14,400,000 records a day and 432,000,000 records every month. For practical purposes, in order to cut down transmission and storage costs, the data-transmission interval is set between 30 seconds to 5 minutes. Even in doing so, analyzing such enormous data using conventional techniques is a mind-boggling task and hence, the technique of data mining is employed to explore useful information from such enormous data. Deviations from the usual trends are of specific interest to the administrators, since such data hold the key for detecting anomalies and implementing corrective measures.

Spatial Data Mining Tasks

Spatial data mining, i.e., extracting useful information from huge amounts of data, is highly relevant to applications in which the tremendous data volumes are involved, thus exceeding human analytical capabilities. Data mining is no longer restricted to the relational databases alone. Spatial data mining is proving to be a promising technique, and holds the key to solving several challenging issues concerning spatial databases. Data Mining offers intelligent functionalities and works well with complicated algorithms, such as, neural networks, rule induction, decision trees, and genetic algorithms and constantly updates its models based on self-learning. Some of the widely accepted and applied methodologies of spatial data mining include Association Rules, Characteristic Rules, Discriminated Rules, Classification, Clustering and Trend Detection.

The following sentences briefly illustrate the aforesaid data mining methodologies.

1. Association Rules: The notion of *association rules* was put forward by Agrawal *et al.* [1993] in a study of mining large transaction databases. The structure of an association rule is: $A \rightarrow B$(c %), where A and B are predicate sets and c% is the confidence level. For instance, an example association rule is: If the city is large, the probability of its being near to a river is 80%.
2. Characteristic Rules: These are rules concerning the characteristics of the mined objects and these are rules formulated by specialists and may not be completely appropriate. An example of a characteristic rule is: A bridge is an object that is

present at the location where a road crosses a river.

3. Discriminatory Rules: These are rules that aim to differentiate between groups of objects by finding features that are close to one cluster and far-off from another. Such a rule is applied to find the differences between cities with high and low unemployment rates.
4. Classification rules: These rules classify a pixel into one of the given set of classes, e. g. water, field, forest. IF population of city = high AND economic power of city = high THEN unemployment of city is classified as " low".
5. Clustering: Clustering involves grouping pixels into similarity classes based on spectral characteristics. An example is to find clusters of cities with similar levels of unemployment.
6. Trend Detection: Trend detection refers to looking for similarity in prototype among the mined objects. An example of trend detection may be as follows: when moving away from Brno, the unemployment rate increases.
7. Sequential Pattern: These are used to detect sequences of events or values, such as stock and share values, business transactions etc.

Applying Data Mining to Fleet Management

Shipping agencies and transportation companies would be interested in the occurrences of some particular kinds of situations. For instance, a company would want to know if a driver follows the stipulated routes, or if the driver observes traffic rules, or if a driver takes a break longer than the allotted time-interval etc. Several factors including efficiency of delivery, attendance rate, observation of traffic rules etc. may be used to judge the driver's performance. The two broad sections of fleet management are operational management and the management of abnormal conditions.

Primarily, data mining and other advanced information techniques are employed in order to improve efficiency of operations. During several occasions, the compensation or financial loss arising out of accidents or delayed deliveries amounts to dozens of millions. Hence, it is imperative that transportation vendors learn from prior operational experiences and take steps to reduce financial loss due to abnormal occurrences.

Data mining is used to locate hidden patterns in the data by calculating the frequency of occurrence among the items while running the permutations and combinations. The main pattern is identified upon detecting a frequency that is higher than the standard frequency of that particular item set. Vehicle behavior is characterized by a large proportion of vehicle records that contain details of speed, location etc. Abnormal conditions like over-speeding or route deviation constitute only a little proportion of the whole. In order to obtain operation statistics and other related operational details the normal vehicle records can be analyzed. Analysis to locate abnormal records like those of illegal parking or over-speeding may be complicated and hence will be studied with many sequential records.

Data Resource

A brief understanding of the data resources is inevitable prior to the application of data mining for record analysis. The following table depicts a sample vehicle record. The fields include car ID, track time, status, speed, directions, coordinates, and dispatch time. This table reveals the sequential characteristic wherein every record is maintained on a chronological or regular time basis. In addition, for each instance the dispatch time is used to identify the vehicle number. The same dispatch time means the same vehicle number. The field values under car ID and dispatch time can be used in conjunction to get information like names of drivers or shipment details.

CarID	Track Time	Status	Speed	Angle	X	Y	Dispatch Time
0007	00:01:09	0	54	326	175905.0658	2634542.6274	00:01:09
0007	00:14:43	0	62	112	166802.7410	2632986.0253	00:01:09
0007	00:28:19	2	0	074	183205.5324	2652531.2527	00:01:09
0007	01:25:30	0	43	351	190947.0462	2660255.8293	00:01:09

0007	02:07:41	65	106	003	200352.7703	2662330.6449	02:07:41
0007	03:13:55	0	80	009	204302.4380	2666640.1632	02:07:41
0007	03:24:31	3	0	267	216812.8626	2683926.0964	02:07:41
…							

Table 1: the example of the vehicle records

Data Pre-processing

Obtaining sequential pattern from the original data is not a straightforward process. For instance, to find driving routes, the vehicle record has to be converted to a meaningful location in coordinates. Then, it can be determined as to where to start out, pass along, and where to arrive at. In doing so even the variation in driving time and speed can be found out. However, lots of valuable time will be wasted if the analytic procedure is executed every time. A better option is to perform data pre-processing initially.

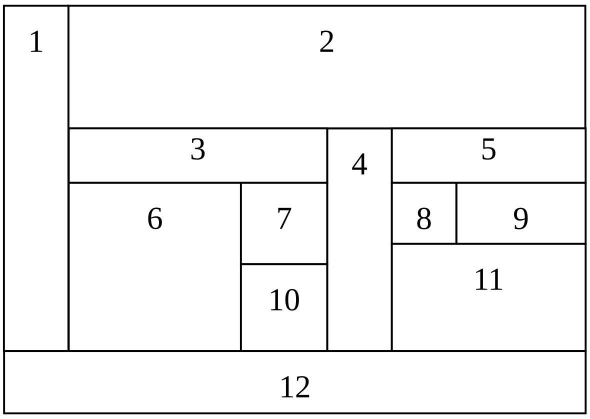

Figure 1: Region Partition

We propose the concept “region partition” wherein each region is divided into several rectangles, with a unique Location Number for each rectangle. Each rectangle is called the Minimum Bounding Rectangle (MBR). The principal advantage of this concept is that it is swift; yet it is not accurate enough. As a matter of fact, this shortfall was conquered later by developing one concept- polygon algorithm.

From Table 2, we can find out the region referring to coordinates of vehicle records and transfer to the corresponding 'Location No.' in advance. Subsequently, using sequential patterns to analyze the data becomes much easier.

No.	Caption	X1	Y1	X2	Y2
2	Feng Chia University	166802.7410	2632986.0253	169813.5124	2638012.4312
3	Wenhua Rd.	183205.5324	2652531.2527	183426.4163	2657784.1105
4	Fengjia Rd.	216812.8626	2683926.0964	216935.8626	2687850.5541
…		…			
10	McDonalds	175455.5245	2606565.5105	178874.3467	2625635.0961
11	Bigen Plaza	206540.6402	2705210.0431	206783.1465	2762346.6323
12	Fushing Rd	193365.6548	2680565.5457	193732.2434	2686733.3541
…		…			

Table 2: Regions and Location No.

Car ID	Track Time	Status	Speed	Angle	No.	Caption	Dispatch Time
0007	00:01:09	0	54	326	2	Feng Chia University	00:01:09
0007	00:14:43	0	62	112	3	Wenhua Rd.	00:01:09
0007	00:28:19	2	0	074	4	Fengjia Rd.	00:01:09
0007	01:25:30	0	43	351	11	Bigen Plaza	00:01:09
0007	02:07:41	65	106	003	10	McDonalds	02:07:41
0007	03:13:55	0	80	009	12	Fushing Rd	02:07:41
0007	03:24:31	3	0	267	12	Fushing Rd	02:07:41
…							

Table 3: transfer to correspondent Location No.

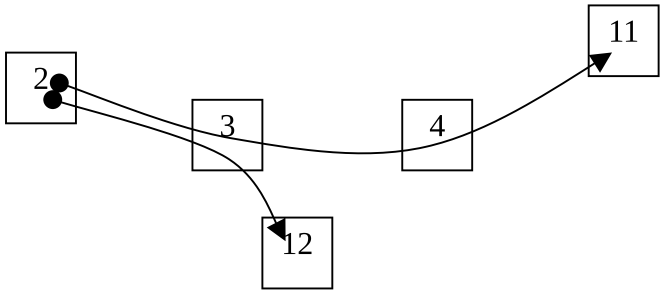

Figure 2: regular routes and deviating routes

Analysis

The application of data mining can lead to immensely useful results and these cannot be elaborated in detail here, due to the nature and scope of this paper. However, in our case plenty of useful information was obtained by employing data mining (Table 3). Figure 2 below, provides one example to explain "Sequential Pattern" in Data Mining. Sequential Pattern was utilized to locate the regular routes (Figure 2: Check Point 2, 3, 4, 11) for the reference of operational analysis. Comparing these regular routes with the current status, the system identifies any deviation (Check Point 2: Location 2, 3, 12) and alerts the administrator. However erroneous judgments are, to a certain extent possible, with sequential pattern. Particularly so, when the inaccuracy exists in GPS itself and in such a case the ratio of errors might be higher.

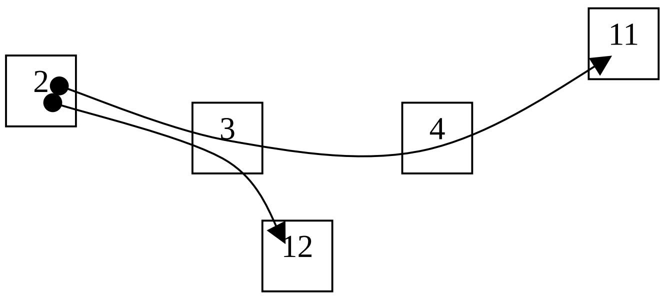

Figure 2: regular routes and deviating routes

For the purpose of comparison some salient features are selected and evaluated. The example above defines many "Check Points" in one route. If the vehicle traverses these check-points, then the vehicle is regarded to follow the driving routes. There are many techniques to set up check points. For instance, while establishing check points on the highway, one check point can be set up on each interchange respectively. By doing so, one erroneous judgment of the deviating route on the highway would not affect the judgment in the whole highway. Moreover, errors by the front and rear locations can be rectified to reduce probability of erroneous judgments.

Additionally, cluster analysis can be used to detect vehicles halting and staying around some place. From the vehicles track records the data corresponding to '0' speed value can be selected, and all legal parking place records, records of places for loading and unloading cargo and records of crossroads(traffic lights) can be eliminated. After this process of eliminating all permitted '0' speed levels, if there still exist some points close enough to form some clusters, this might refer to some place where most vehicle drivers prefer to stop their vehicles and stay around. Those are atypical parking places that can be mined or extracted out from vehicle track record database and this can also be used to detect whether drivers involve in illegal matters at these sites.

Conclusions

This paper has presented an approach to apply the technique of data mining for commercial vehicle operation and management processes and this study is intended to serve as a reference for developing advanced fleet management systems. The Fleet Management Systems have come a long way from passive monitoring systems to active tracking systems. However, these fleet management systems lack intelligence and users are looking forward to a smarter system. Hence this research embarked on data mining, thus developing a system that can track vehicles with artificial intelligence. Using such a system that can detect vehicle behavior in advance, various abnormal conditions can be avoided and traffic accidents minimized.

In addition to tracking vehicles, the fleet management system proposed herein can also be extended to track different targets such as persons including elders, children and animals. Those records can be analyzed using data mining techniques to find out interesting patterns. However due to the lower calculation speeds data mining today is rather slow and hence fleet management systems can only analyze records after incident-occurrences, and can't analyze vehicle status in real time. Future systems must integrate GPS with real time video capture, as used in some of the video mapping systems. Even though such a facility will provide more valuable information, it will lead to data accumulation and hence make data Mining a more challenging task.

References

1. Jiawei Han, Micheline Kamber, "Data Mining Concept and Techniques". Morgan Kaufmann Publisheres 2001, ISBN 1-55860-489-8
2. R. Agrawal, T. Imielinski, and A. Swami. Mining Association Rules Between Sets of Items in Large Databases. In Proc. 1993 ACM-SIGMOD Int. Conf. Management of Data, pp. 207--216, Washington, D.C., May 1993.
3. Michael J.A.Berry, and Gordon S. Linoff, "Data Mining Techniques: for marketing, sales, and customer support". Taiwan: SuperPoll.net, Inc. January 2001. ISBN 957-8675-75-5
4. Jun Chen, "Voronoi Dynamic Spatial Data Model". China: Surveying and Drawing Publisheres, August 2002. ISBN 7-5030-1119-X
5. Kai-Chang Di, "Spatial Data Mining and Knowledge Discovery", China: Wuhan University Publisheres, December 2001. ISBN 7-307-03270-8
6. He-Hai Wu, Jian-Ya Gong, "Geographic Information Systems Spatial Data Structure and Process Technology". China: Surveying and Drawing Publisheres , April 1997. ISBN 7-5030-0937-3
7. Peng Hu, Xing-Yuan Huang, Yi-Xin Hua, "Geographic Information Systems Course". China: Wuhan University Publisheres , February 2002. ISBN

7-307-03432-8

8. Rong-Chung Chen, Yu-Chen Lin, “A Study of Clustering Technology”. Department of Information Management Chaoyang University of Technology Thesis for the Degree of Master, August 2002.

9. Petr Kuba, “Data Structures for Spatial Data Mining”, Department of Computer Science, Faculty of Informatics Masaryk University Brno, Czech Republic, September 4, 2001

10. Marc van Kreveld, Jurg Nievergelt, Thomas Roos, and Peter Widmayer, “Algorithmic Foundations of Geographic Information Systems”. Tokyo: Springer-Verlag, 1997. ISBN 3-540-63818-0