ALGORITHM FOR DETECTION OF HOT SPOTS OF TRAFFIC THROUGH ANALYSIS OF GPS DATA

Thesis submitted in partial fulfillment of the requirements for the award of degree of

Master of Engineering
in
Computer Science and Engineering

By:
Jitesh Pati Tripathi
(800832006)

Under the supervision of:
Dr. Deepak Garg
Assistant Professor, CSED

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT
THAPAR UNIVERSITY
PATIALA – 147004

JUNE - 2010

Certificate

I hereby certify that the work which is being presented in the thesis entitled, **"Algorithm for Detection of Hot Spots of Traffic Through Analysis of GPS Data"**, in partial fulfillment of the requirements for the award of degree of Master of Engineering in Computer Science and Engineering submitted in Computer Science and Engineering Department of Thapar University, Patiala, is an authentic record of my own work carried out under the supervision of Dr. Deepak Garg and refers other researcher's works which are duly listed in the reference section.

The matter presented in this thesis has not been submitted for the award of any other degree of this or any other university.

(Jitesh Pati Tripathi)

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.

(Dr. Deepak Garg)
Assistant Professor
Computer Science and Engineering Department

Countersigned by

02/07/10
(Dr. Rajesh Bhatia)
Head
Computer Science and Engineering, Department
Thapar University, Patiala

(Dr. R. K. Sharma) 13/7/1·
Dean (Academic Affairs)
Thapar University,
Patiala

Acknowledgement

No volume of words is enough to express my gratitude towards my guide **Dr. Deepak Garg**, Department of Computer Science and Engineering, Thapar University, Patiala. He has helped me to explore this vast topic in an organized manner and provided me with all the ideas on how to work towards a research-oriented venture.

I am also thankful to **Dr. Rajesh Bhatia,** Head of Department, Computer Science and Engineering Department and **Dr. Inderveer Chana**, P.G. Coordinator, for the motivation and inspiration that triggered me for the thesis work.

I would also like to thank the staff members and my colleagues who were always there at the need of the hour and provided with all the help and facilities.

Most importantly, I would like to thank my parents and the almighty for showing me the right direction out of the blue, to help me stay calm in the oddest of the times and keep moving even at times when there was no hope.

Jitesh Pati Tripathi
(800832006)

Abstract

Traffic congestion continues to be one of the major problems in various transportation systems. The ever increasing demand for mobility in urban centers has resulted in increased traffic congestion and a multitude of problems associated with it. People are especially concerned about those areas of the city which are regularly congested, because these areas could magnify the impact area and the duration of congestion. A prior knowledge about these areas and better understanding of impact of these areas could help into more appropriate traffic management strategies.

This report present the method of detecting these areas ,which we call as 'Hot-Spots' in terms of road traffic and analyses the impact of these 'Hot-Spots' on traffic, both in terms of time and space. The Global Positioning System (GPS) data is used for this objective, as GPS has become a reliable, accurate and economically feasible positioning technology for probe vehicle. In the report, we present algorithms for the detection of the 'Hot-Spots', based on the speed of the probe vehicle and clustering (Fuzzy c-means algorithm) algorithm.

.

Contents Page No.

List of Figures

Figure no.	Title	Page no.

CHAPTER 1

INTRODUCTION

1.1 Problem Statement

Traffic congestion continues to be one of the major problems in various transportation systems. The dramatic increase in traffic volume worldwide is leading to massive congestions causing various social, environmental and economic problems.

Congestion are often caused by or made worse by traffic incidents. A sudden traffic surge immediately after special events or some special location (‘Hot Spots’) in the city can create substantial traffic congestion in the area related with these incidents or Hot Spots.

Hot-Spots and road incident induced traffic congestion provide real threats to the mobility and safety of our daily travel. Congestion may be alleviated by providing timely and accurate traffic information so that motorists can avoid congested routes by using alternative routes or changing their departure times. In general, the public tends to think more in terms of travel time rather than volume in evaluating the quality of their trips. Accordingly, hot spot information is a key element for avoiding congestion as well as for making informed driving route decision.

Therefore timely and accurate detection of road incident and hot-spots is an essential part of any successful advanced traffic management system. If any incidents or Hot Spots are detected quickly, they can be cleared swiftly, resulting in less traffic congestion, and then more appropriate traffic management strategies can be applied to provide better service to the road users. So it is desired that the Hot-Spots related traffic performance to be measured so that the traffic flow can be improved.

1.2 Objective

The GPS data provides sufficient information for identifying general traffic patterns, such as the average speed on a specified road, abnormal behavior of a vehicle like unusual deceleration or change in direction. Such real-time information may be able to infer the occurrence of an accident. GPS has become a reliable, accurate, economically feasible and the most recent positioning technology to be used for travel time data collection [1].

In this work, the aim is to identify the different 'hot-spots' of the city.

1.3 Scope of work

The study involves the detection of the 'hot-spots' and their impact on the traffic, as the prior knowledge of these areas is very important for better traffic management strategies. GPS data is used for this purpose.

The Thesis report is organized in seven chapters. In chapter 2, we have described the definition of 'hot-spots'& GPS system. Chapter 3 is Literature review describe the previous work done for identifying traffic incidents and clustering algorithms. Chapter 4 describes the collection of data and mapping of data on road map of Mumbai. Chapter 5 explains about the algorithm for detection of congested areas and results. Chapter 6 described the clustering (Fuzzy c-means algorithm) algorithms and Chapter 7 contains the conclusion and future scope of the Thesis report.

CHAPTER 2

HOT-SPOTS AND GPS SYSTEM

2.1 Hot-Spots

In any city 'hot-spots' are the locations or those area of the city, which are regularly congested and where the flow of traffic is very slow that is traffic is moving very slowly or traffic is completely stopped. These areas could increase the severity of congestion.

Hot-spots can be of two types; regularly and occasionally.

Multiplexes, Commercial buildings, Hospitals, Schools/Colleges etc. are prime examples of regular hot-spots. Occasionally hot-spots are the areas where any incident has taken place. An incident is "an unexpected event that temporarily disrupts the flow of traffic on a segment of a roadway". For example: area where an accident has taken place, VIP passing area.

Hot-spots and motorway incident are responsible for a significant proportion of delays and costs to the motoring public. Hot-spots could increase the severity of recurrent congestion in terms of the impact area and duration. Hot-spots detection is the process that brings hot-spots to the attention of those responsible for maintaining traffic flow and safe operations. The impact of hot-spots can be reduced through a variety of actions, including the use of traffic information, ramp restrictions or closure, and advice on alternative routes.

Therefore, faster detection of these areas is of extreme importance for maintaining a higher level service for road users, also enables agencies to respond more quickly in removing the problem, to warn the oncoming traffic, and to reduce the danger of secondary incidents.

2.2 GPS System [1]

Fig 2.1: GPS System

The Global Positioning System (GPS) is a U.S.-owned utility that provides users with positioning, navigation, and timing (PNT) services. This system consists of three segments: the space segment, the control segment, and the user segment. The U.S. Air Force develops, maintains, and operates the space and control segments.

- The space segment consists of a nominal constellation of 24 operating satellites that transmit one-way signals that give the current GPS satellite position and time.
- The control segment consists of worldwide monitor and control stations that maintain the satellites in their proper orbits through occasional command maneuvers, and adjust the satellite clocks. It tracks the GPS satellites, uploads updated navigational data, and maintains health and status of the satellite constellation.

1, Article 2.2 referenced from http://www.gps.gov/system/gps [1]

- The user segment consists of the GPS receiver equipment, which receives the signals from the GPS satellites and uses the transmitted information to calculate the user's three-dimensional position and time.

2.2.1 GPS Services

GPS satellites provide service to civilian and military users. The civilian service is freely available to all users on a continuous, worldwide basis. The military service is available to U.S. and allied armed forces as well as approved Government agencies.

A variety of GPS augmentation systems and techniques are available to enhance system performance to meet specific user requirements. These improve signal availability, accuracy, and integrity, allowing even better performance than is possible using the basic GPS civilian service.

The outstanding performance of GPS over many years has earned the confidence of millions of civil users worldwide. It has proven its dependability in the past and promises to be of benefit to users, throughout the world, far into the future.

2.2.2 The Future of GPS

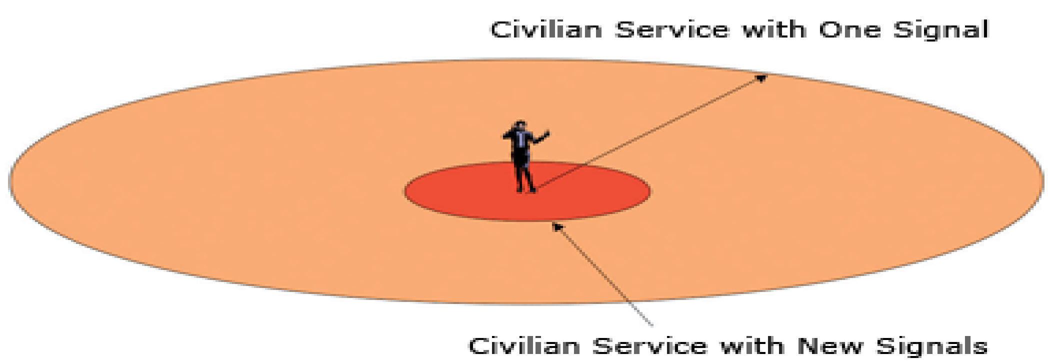

Fig 2.2: Stand-Alone GPS Notional Horizontal Performance with New Signals

The United States is committed to an extensive modernization program, including the implementation of a second and a third civil signal on GPS satellites. The second civil signal will improve the accuracy of the civilian service and supports some safety-of-life applications. The third signal will further enhance civilian capability and is primarily designed for safety-of-life applications, such as aviation.

The figure on the right depicts the improvement of the quality of service with the additional civilian signals.

2.2.3 Positioning, Navigation, and Timing Policy

U.S. law and policy on GPS emphasize continuity of service, open access to civil signals, and technological leadership. In 1996, the United States issued a national policy statement on the management and use of space-based positioning, navigation and timing services, which include GPS and augmentations. It underscored the dual use (civilian-military) nature of GPS and established a joint civil-military national management structure to oversee its operation.

U.S. policy was expanded in 2004 in response to changing international conditions and the incredible growth in the types and complexities of GPS applications. This policy reaffirms the United States commitment to provide reliable civil space-based positioning, navigation, and timing services through GPS on a continuous, worldwide basis -- freely available to all. The policy also calls for improving the performance of GPS and cooperating with other nations.

2.2.4 GPS Augmentations

To meet the specific user requirements for positioning, navigation, and timing (PNT), a number of augmentations to the Global Positioning System (GPS) are available. An augmentation is any system that aids GPS by providing accuracy, integrity, reliability, availability, or any other improvement to positioning, navigation, and timing that is not inherently part of GPS itself. Such augmentations include, but are not limited to:

- **Nationwide Differential GPS System (NDGPS):** The NDGPS is a ground-based augmentation system operated and maintained by the Federal Railroad

Administration, U.S. Coast Guard, and Federal Highway Administration, that provides increased accuracy and integrity of the GPS to users on land and water. Modernization efforts include the High Accuracy NDGPS (HA-NDGPS) system, currently under development, to enhance the performance and provide 10 to 15 centimeter accuracy throughout the coverage area. NDGPS is built to international standards, and over 50 countries around the world have implemented similar systems.

- **Wide Area Augmentation System (WAAS):** The WAAS, a satellite-based augmentation system operated by the U.S. Federal Aviation Administration (FAA), provides aircraft navigation for all phases of flight. Today, these capabilities are broadly used in other applications because their GPS-like signals can be processed by simple receivers without additional equipment. Using International Civil Aviation Organization (ICAO) standards, the FAA continues to work with other States to provide seamless services to all users in any region. Other ICAO standard space-based augmentation systems include: Europe's European Geostationary Navigation Overlay System (EGNOS), India's GPS and Geo-Augmented Navigation System (GAGAN), and Japan's Multifunction Transport Satellite (MTSAT) Satellite Augmentation System (MSAS). All of these international implementations are based on GPS. The FAA will improve the WAAS to take advantage of the future GPS safety-of-life signal and provide better performance and promote global adoption of these new capabilities.
- **Continuously Operating Reference Station (CORS):** The U.S. CORS network, which is managed by the National Oceanic & Atmospheric Administration, archives and distributes GPS data for precision positioning and atmospheric modeling applications mainly through post-processing. CORS is being modernized to support real-time users.
- **Global Differential GPS (GDGPS):** GDGPS is a high accuracy GPS augmentation system, developed by the Jet Propulsion Laboratory (JPL) to support the real-time positioning, timing, and orbit determination requirements of the U.S. National Aeronautics and Space Administration (NASA) science missions. Future NASA plans include using the Tracking and Data Relay Satellite

System (TDRSS) to disseminate via satellite a real-time differential correction message. This system is referred to as the TDRSS Augmentation Service Satellites (TASS).

- **International GNSS Service (IGS):** IGS is a network of over 350 GPS monitoring stations from 200 contributing organizations in 80 countries. Its mission is to provide the highest quality data and products as the standard for Global Navigation Satellite Systems (GNSS) in support of Earth science research, multidisciplinary applications, and education, as well as to facilitate other applications benefiting society. Approximately 100 IGS stations transmit their tracking data within one hour of collection.

There are other augmentation systems available worldwide, both government and commercial. These systems use differential, static, or real-time techniques

2.2.5 U.S. Policy on International Cooperation

The importance that all global navigation satellite systems and their augmentations be compatible with GPS.

The agreement in 2004 between the United States and the European Union (E.U.) on GPS and Galileo recognized the benefits of interoperable systems. The parties agreed to pursue a common, open, civil signal on both Galileo and future GPS satellites, in addition to ongoing cooperation on the GPS-based EGNOS augmentation system.

The United States has a long cooperative relationship with Japan on GPS. In addition to the Multifunction Transport Satellite (MTSAT) Satellite Augmentation System (MSAS), the parties are working towards developing a GPS-compatible regional satellite "mini-" constellation known as the Quasi Zenith Satellite System (QZSS).

The United States is also consulting closely with India on its development of its GAGAN space-based augmentation system, and with the Russian Federation on compatibility and interoperability between GPS and Russia's satellite navigation system, GLONASS.

The U.S. Department of Defense also cooperates with numerous countries to ensure that GPS provides military space-based PNT service and interoperable user equipment to its coalition partners around the world.

Space-based PNT services must serve global users with transparent interfaces and standards. The U.S. policy is to provide space-based PNT services on a continuous worldwide basis, freely available to all for civil, commercial, and scientific uses, and provide open, free access, to information necessary to develop and build equipment to use these services.

2.2.6 Applications of GPS System

2.2.6.1 Timing

In addition to longitude, latitude, and altitude, the Global Positioning System (GPS) provides a critical fourth dimension – time. Each GPS satellite contains multiple atomic clocks that contribute very precise time data to the GPS signals. GPS receivers decode these signals, effectively synchronizing each receiver to the atomic clocks. This enables users to determine the time to within 100 billionths of a second, without the cost of owning and operating atomic clocks.

Precise time is crucial to a variety of economic activities around the world. Communication systems, electrical power grids, and financial networks all rely on precision timing for synchronization and operational efficiency. The free availability of GPS time has enabled cost savings for companies that depend on precise time and has led to significant advances in capability.

For example, wireless telephone and data networks use GPS time to keep all of their base stations in perfect synchronization. This allows mobile handsets to share limited radio spectrum more efficiently. Similarly, digital broadcast radio services use GPS time to ensure that the bits from all radio stations arrive at receivers in lockstep. This allows listeners to tune between stations with a minimum of delay.

Companies worldwide use GPS to time-stamp business transactions, providing a consistent and accurate way to maintain records and ensure their traceability. Major investment banks use GPS to synchronize their network computers located around the world. Large and small businesses are turning to automated systems that can track, update, and manage multiple transactions made by a global network of customers, and these require accurate timing information available through GPS.

The U.S. Federal Aviation Administration (FAA) uses GPS to synchronize reporting of hazardous weather from its 45 Terminal Doppler Weather Radars located throughout the United States.

Instrumentation is another application that requires precise timing. Distributed networks of instruments that must work together to precisely measure common events require timing sources that can guarantee accuracy at several points. GPS-based timing works exceptionally well for any application in which precise timing is required by devices that are dispersed over wide geographic areas. For example, integration of GPS time into seismic monitoring networks enables researchers to quickly locate the epicenters of earthquakes and other seismic events.

Power companies and utilities have fundamental requirements for time and frequency to enable efficient power transmission and distribution. Repeated power blackouts have demonstrated to power companies the need for improved time synchronization throughout the power grid. Analyses of these blackouts have led many companies to place GPS-based time synchronization devices in power plants and substations. By analyzing the precise timing of an electrical anomaly as it propagates through a grid, engineers can trace back the exact location of a power line break.

Some users, such as national laboratories, require the time at a higher level of precision than GPS provides. These users routinely use GPS satellites not for direct time acquisition, but for communication of high-precision time over long distances. By simultaneously receiving the same GPS signal in two places and comparing the results, the atomic clock time at one location can be communicated to the other. National laboratories around the world use this "common view" technique to compare their time scales and establish Coordinated Universal Time (UTC). They use the same technique to disseminate their time scales to their own nations.

New applications of GPS timing technology appear every day. Hollywood studios are incorporating GPS in their movie slates, allowing for unparalleled control of audio and video data, as well as multi-camera sequencing. The ultimate applications for GPS, like the time it measures, are limitless.

As GPS becomes modernized, further benefits await users. The addition of the second and third civilian GPS signals will increase the accuracy and reliability of GPS time, which will remain free and available to the entire world.

2.2.6.2 Roads & Highways

It is estimated that delays from congestion on highways, streets, and transit systems throughout the world result in productivity losses in the hundreds of billions of dollars annually. Other negative effects of congestion include property damage, personal injuries, increased air pollution, and inefficient fuel consumption.

The availability and accuracy of the Global Positioning System (GPS) offers increased efficiencies and safety for vehicles using highways, streets, and mass transit systems. Many of the problems associated with the routing and dispatch of commercial vehicles is significantly reduced or eliminated with the help of GPS. This is also true for the management of mass transit systems, road maintenance crews, and emergency vehicles,

GPS enables automatic vehicle location and in-vehicle navigation systems that are widely used throughout the world today. By combining GPS position technology with systems that can display geographic information or with systems that can automatically transmit data to display screens or computers, a new dimension in surface transportation is realized.

A geographic information system (GIS) stores, analyzes, and displays geographically referenced information provided in large part by GPS. Today GIS is used to monitor vehicle location, making possible effective strategies that can keep transit vehicles on schedule and inform passengers of precise arrival times. Mass transit systems use this capability to track rail, bus, and other services to improve on-time performance.

Many new capabilities are made possible with the help of GPS. Instant car pools are feasible since people desiring a ride can be instantly matched with a vehicle in a nearby area.

Using GPS technology to help track and forecast the movement of freight has made a logistical revolution, including an application known as time-definite delivery. In time-definite delivery, trucking companies use GPS for tracking to guarantee delivery and pickup at the time promised, whether over short distances or across time zones. When an order comes in, a dispatcher punches a computer function, and a list of trucks appears on the screen, displaying a full array of detailed information on the status of each of them. If a truck is running late or strays off route, an alert is sent to the dispatcher.

Many nations use GPS to help survey their road and highway networks, by identifying the location of features on, near, or adjacent to the road networks. These include service stations, maintenance and emergency services and supplies, entry and exit ramps, damage to the road system, etc. The information serves as an input to the GIS data gathering process. This database of knowledge helps transportation agencies to reduce maintenance and service costs and enhances the safety of drivers using the roads.

Research is underway to provide warnings to drivers of potential critical situations, such as traffic violations or crashes. Additional research is being conducted to examine the potential for minimal vehicle control when there is a clear need for action, such as the pre-deployment of air bags. The position information provided by GPS is an integral part of this research.

GPS is an essential element in the future of Intelligent Transportation Systems (ITS). ITS encompasses a broad range of communications-based information and electronics technologies. Research is being conducted in the area of advanced driver assistance systems, which include road departure and lane change collision avoidance systems. These systems need to estimate the position of a vehicle relative to lane and road edge with an accuracy of 10 centimeters.

With the continuous modernization of GPS, one can expect even more effective systems for crash prevention, distress alerts and position notification, electronic mapping, and in-vehicle navigation with audible instructions.

2.2.6.3 Space

2.2.6.3.1 Earth Orbit

The Global Positioning System (GPS) is revolutionizing and revitalizing the way nations operate in space, from guidance systems for crewed vehicles to the management, tracking, and control of communication satellite constellations, to monitoring the Earth from space. Benefits of using GPS include:

- Navigation solutions -- providing high precision orbit determination, and minimum ground control crews, with existing space-qualified GPS units.
- Attitude solutions -- replacing high cost on-board attitude sensors with low-cost multiple GPS antennae and specialized algorithms.
- Timing solutions -- replacing expensive spacecraft atomic clocks with low-cost, precise time GPS receivers.
- Constellation control -- providing single point-of-contact to control for the orbit maintenance of large numbers of space vehicles such as telecommunication satellites.
- Formation flying -- allowing precision satellite formations with minimal intervention from ground crews.
- Virtual platforms -- providing automatic "station-keeping" and relative position services for advanced science tracking maneuvers such as interferometry.

- Launch vehicle tracking -- replacing or augmenting tracking radars with higher precision, lower-cost GPS units for range safety and autonomous flight termination.

2.2.6.3.2 The Moon, Mars, and Beyond

The U.S. vision for space exploration, being implemented by the National Aeronautics and Space Administration (NASA), includes developing innovative technologies, knowledge, and infrastructures for returning to the Moon and preparing the way for future human missions to Mars and beyond. The vision will stimulate new research that will literally become the final frontier in navigation. Drawing on the experience with GPS, one could imagine creating a GPS-like network of satellites around the Moon and Mars. A Lunar or Martian network could provide an integrated communications and navigation infrastructure to support exploration and science missions both in lunar orbit and on the surface of the Moon and Mars.

NASA is also studying the utility of placing GPS-like beacons on satellites destined for the Sun-Earth Lagrangian points. Geodetic reference points could be established at these locations to support the future exploration of the Solar System.

2.2.6.4 Agriculture

The development and implementation of precision agriculture or site-specific farming has been made possible by combining the Global Positioning System (GPS) and geographic information systems (GIS). These technologies enable the coupling of real-time data collection with accurate position information, leading to the efficient manipulation and analysis of large amounts of geospatial data. GPS-based applications in precision farming are being used for farm planning, field mapping, soil sampling, tractor guidance, crop scouting, variable rate applications, and yield mapping. GPS allows farmers to work during low visibility field conditions such as rain, dust, fog, and darkness.

In the past, it was difficult for farmers to correlate production techniques and crop yields with land variability. This limited their ability to develop the most effective soil/plant treatment strategies that could have enhanced their production. Today, more precise application of pesticides, herbicides, and fertilizers, and better control of the dispersion of those chemicals are possible through precision agriculture, thus reducing expenses, producing a higher yield, and creating a more environmentally friendly farm.

Precision agriculture is now changing the way farmers and agribusinesses view the land from which they reap their profits. Precision agriculture is about collecting timely geospatial information on soil-plant-animal requirements and prescribing and applying site-specific treatments to increase agricultural production and protect the environment. Where farmers may have once treated their fields uniformly, they are now seeing benefits from micromanaging their fields. Precision agriculture is gaining in popularity largely due to the introduction of high technology tools into the agricultural community that are more accurate, cost effective, and user friendly. Many of the new innovations rely on the integration of on-board computers, data collection sensors, and GPS time and position reference systems.

Many believe that the benefits of precision agriculture can only be realized on large farms with huge capital investments and experience with information technologies. Such is not the case. There are inexpensive and easy-to-use methods and techniques that can be developed for use by all farmers. Through the use of GPS, GIS, and remote sensing, information needed for improving land and water use can be collected. Farmers can achieve additional benefits by combining better utilization of fertilizers and other soil amendments, determining the economic threshold for treating pest and weed infestations, and protecting the natural resources for future use.

GPS equipment manufacturers have developed several tools to help farmers and agribusinesses become more productive and efficient in their precision farming activities. Today, many farmers use GPS-derived products to enhance operations in their farming businesses. Location information is collected by GPS receivers for mapping field boundaries, roads, irrigation systems, and problem areas in crops such as weeds or disease. The accuracy of GPS allows farmers to create farm maps with precise acreage for field areas, road locations and distances between points of interest. GPS allows farmers to accurately navigate to specific locations in the field, year after year, to collect soil samples or monitor crop conditions.

Crop advisors use rugged data collection devices with GPS for accurate positioning to map pest, insect, and weed infestations in the field. Pest problem areas in crops can be pinpointed and mapped for future management decisions and input recommendations. The same field data can also be used by aircraft sprayers, enabling accurate swathing of fields without use of human “flaggers” to guide them. Crop dusters equipped with GPS are able to fly accurate swaths over the field, applying chemicals only where needed, minimizing chemical drift, reducing the amount of chemicals needed, thereby benefiting the environment. GPS also allows pilots to provide farmers with accurate maps.

Farmers and agriculture service providers can expect even further improvements as GPS continues to modernize. In addition to the current civilian service provided by GPS, the United States is committed to implementing a second and a third civil signal on GPS satellites. The first satellite with the second civilian signal was launched in 2005. The new signals will enhance both the quality and efficiency of agricultural operations in the future.

2.2.6.5 Rail

Rail systems in many parts of the world use the Global Positioning System (GPS) in combination with various sensors, computers, and communication systems to improve safety, security, and operational effectiveness. These technologies help to reduce accidents, delays, operating costs, and dangerous emissions, while increasing track capacity, customer satisfaction, and cost effectiveness. Integral to the efficient operation of rail systems is the requirement for accurate, real-time position information of locomotives, rail cars, maintenance-of-way vehicles, and wayside equipment.

Ensuring high levels of safety, improving the efficiency of rail operations, and expanding system capacity are all key objectives of today’s railroad industry. Unlike most other modes of transportation, there is little flexibility in managing rail traffic. Most rail systems are comprised of long stretches of a single set of tracks. Trains bound for thousands of destinations must simultaneously share the use of these single line tracks.

Precise knowledge of where a train is located is essential to prevent collisions, maintain smooth flow of traffic, and minimize costly delays due to waiting for clearance for track use. Only the skill of the crews, accurate timing, a dynamic dispatching capability, and a critical array of "meet and passes" locations on short stretches of parallel tracks, allow rail dispatchers to guide their trains safely through. It is therefore critical for safety and efficiency reasons to know the position and performance of these trains both individually and system-wide.

GPS also contributes to dependable scheduling through train location awareness, enhancing connectivity with other modes of transportation, such as rail station to airport transfers.

An enhancement to the basic GPS signal known as Differential GPS (DGPS) improves accuracy and safety within its coverage areas. The enhanced position information enables the dispatcher to determine on which of two parallel tracks a train is located. When coupled with other location and navigation devices to account for time in tunnels, behind hills, and other obstructions, DGPS can provide a reliable and accurate position-locating capability for rail traffic management systems.

Differential GPS is an essential element of the Positive Train Control (PTC) concept being adopted in many parts of the world. This concept involves providing precise railroad position information to sophisticated command and control systems to produce the best operating plan to include varying train speed, re-routing traffic, and safely moving maintenance crews onto and off tracks.

A PTC system can track the location and speed of a train more accurately than was previously possible, providing train movement information to rail management personnel who can then enforce speeds and limits of authority, as necessary. By providing better tracking of train location and speed, PTC increases operational efficiency, allows higher track capacity, enhances crew, passenger, and cargo safety, and also results in a safer environment for personnel working on the track.

Differential GPS can also aid in surveying and mapping track structure for maintenance and future system planning. By using DGPS, one can precisely locate mileposts, signal masts, switch points, bridges, road crossings, signal equipment, etc. GPS can satisfy the high level of accuracy needed for operation in terminal areas and rail yards, where dozens of tracks may run in parallel.

Finally, with the modernization of GPS, rail operators can look forward to providing better service. In addition to the current GPS civilian service, the United States is committed to implementing two additional civilian signals. Access to the new signals will mean increased accuracy, more availability, and better integrity for all users.

CHAPTER 3

LITERATURE REVIEW

J-S Yang focused on the study of the arterial travel time prediction using the Kalman filtering and estimation technique. He used travel time as a performance measure due to the following reasons:

(1) It is the most common way that users measure the quality of their trip.

(2) It is a variable that can be directly measured; and

(3) It is a simple measure to use for traffic monitoring.

He studied how easy it is to exit the area where traffic is very congested regularly? And how much does that "ease of movement" vary after the special events? The Global Positioning System (GPS) test vehicle technique is used to collect after events travel time data. Based on the real-time data collected, a discrete-time Kalman filter is then applied to predict travel time exiting the area under study [2].

Shoaib Kamran presented a multilevel approach for detecting traffic incident causing congestion on major roads. It incorporates algorithms to detect unusual traffic patterns and vehicle behaviors on different road segments by utilizing the real-time GPS data obtained from vehicles [3]. The incident detection process involves two phases:

1) Identification of road segments where abnormal traffic pattern is observed and further divides the 'abnormal segments' into smaller segments in order to isolate the potential incident area.

2) Detection of any occurrence of abnormal behavior within the 'abnormal' road section identified in phase 1.

The strength of such approach lies in analyzing vehicle data specific to the identified road segment. In this way, the processing of vast data is avoided which is an essential requirement for the better performance of such complex systems.

C. Basnayake presented a study in which a transit vehicle fleet was used as the probe system for traffic incident detection. Since transit vehicles operate for a primary purpose other than traffic monitoring, the data they provide may be biased in that they may not represent the behavior of a majority of vehicles in the traffic flow. For instance, transit vehicles have to stop at designated stops and hence the reported travel times may contain dwelling times and lost times in approaching and departing the transit stop zones. This paper addresses such issues and proposes several algorithms to modify transit data for incident detection purposes. The detection algorithm is based on two intuitive characteristics of traffic flow. First characteristic is the time taken for a vehicle to travel through a street segment, referred to as the travel time in this paper. The second characteristic is a measure of increased interactions between vehicles after the onset of congestion created by incidents, known as acceleration noise [4].

Y. Li and M. McDonald presented a probe-vehicle-based algorithm to detect incidents on motorways (A broad highway designed for high-speed traffic). The algorithm is based on a bivariate analysis model (BEAM) using two variables: the average travel times of probe vehicles, and the travel time differences between adjacent time intervals. The premise of the model is that link travel times increase more rapidly as a result of a change in capacity (i.e. when an incident occurs) than as a result of a change in demand. The statistical principles of bivariate analysis have been used to study the relationships between the two variables in incident and non-incident conditions [5].

Yuv-Horng Wen and Tsu- Tian Lee presented a systematic process, combining traffic forecasting and data mining models for traffic information systems. Clustering model was developed for mining traffic flow-speed-occupancy relationships, then to extrapolate traffic information. The hybrid grey-based recurrent neural network (G-RNN) was developed for traffic parameter forecasting. Study results were shown that the G-RNN model is capable of predicting traffic parameters with a high degree of accuracy [6].

K. Ozbay, and P. Kachroo presented a new method, using data mining to identify automatically freeway incidents. As a component of a real-time traffic adaptive control system for signal control, the algorithm feeds an incident report to the system's optimization manager, which uses the information to determine the appropriate signal control strategy. Offline tests were conducted to substantiate the performance of the proposed incident detection algorithm based on simulated data [7].

Traffic flow is usually characterized by randomness and uncertainty. Fuzzy logic is known to be well suited for modeling and control such problems. Applications of fuzzy logic in traffic signal control have been made.

The first attempt made to design Fuzzy Traffic Controller was in 90s by Pappis and Mamdani. After that Niittymaki, Kikuchi, Chui and other researchers developed different algorithms and logic controllers to normalize traffic flow [8,10].

3.1 Data Clustering Algorithms [11]

Clustering algorithms are used extensively not only to organize and categorize data, but are also useful for data compression and model construction. This describes some of the online clustering algorithms that can be realized by unsupervised learning neural networks. This chapter introduces some of the most representative off-line clustering techniques frequently used in conjunction with radial basis function networks and fuzzy modeling: (hard) C-means (or K-means) clustering, fuzzy C-means clustering.

Clustering partitions a data set into several groups such that the similarity within a group is larger than that among groups. Achieving such a partitioning requires a similarity metrics that takes two input vectors and returns a value reflecting their similarity. Since most similarity metrics are sensitive to the ranges of elements in the input vectors, each of the input variables must be normalized to within, say, the unit interval [0, 1]. Hence, the rest of this chapter assumes that data set under consideration has already been normalized to be within the unit hypercube.

Clustering techniques are used in conjunction with radial basis function networks or fuzzy modeling primarily to determine initial locations for radial basis functions or fuzzy if-then rules. For this purpose, clustering techniques are validated on the basis of the following assumptions:

1. Similar inputs to the target system to be modeled should produce similar outputs.

2. These similar input-output pairs are bundled into clusters in the training data set.

Assumption 1 states that the target system to be modeled is a smooth input- output mapping; this is generally true for real-world systems. Assumption 2 requires the data set to conform to some specific type of distribution; however, this is not always true.

Therefore, clustering techniques used for structure identification in neural or fuzzy modeling are highly heuristic, and finding a data set to which clustering techniques cannot be applied satisfactorily is not uncommon.

3.1.1 K-Means Clustering

The K-means clustering also known as C-means clustering has been applied to a variety of areas, including image and speech data compression data preprocessing for system modeling using radial basis function networks and task decomposition in heterogeneous neural network architectures.

The K-means algorithm partitions a collection of n vector *Xj*, j = 1, ..., n, into c groups *Gi*, i = 1,... ,c, and finds a cluster center in each group such that a cost function (or an objection function) of dissimilarity (or distance) measure is minimized. When the Euclidean distance is chosen as the dissimilarity measure between a vector x& in group j and the corresponding cluster center Ci, the cost function can be defined by

$$J = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c} \left(\sum_{k,\, x_k \in G_i} \|\mathbf{x}_k - \mathbf{c}_i\|^2 \right) \tag{3.1}$$

where $J_i = \sum_{k,\, x_k \in G_i} \|\mathbf{x}_k - \mathbf{c}_i\|^2$ is the cost function within group i. Thus, the value of *Ji* depends on the geometrical properties of *Gi* and the location of *Ci*.

In general, a generic distance function d(*Xk*, *Ci*) can be applied for vector *Xk* in group i; the corresponding overall cost function is thus expressed as

$$J = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c} \left(\sum_{k,\, x_k \in G_i} d(\mathbf{x}_k - \mathbf{c}_i) \right) \tag{3.2}$$

For simplicity, the Euclidean distance is used as the dissimilarity measure and the overall cost function is expressed as in Equation (3.1).

The partitioned groups are typically defined by an c x n binary membership matrix U, where the element *Uij* is 1 if the *j*th data point *Xj*, belongs to group i, and 0 otherwise.

Once the cluster centers *Ci* are fixed, the minimizing *Uij* for Equation (3.1) can be derived as follows:

$$u_{ij} = \begin{cases} 1 \text{ if } \|\mathbf{x}_j - \mathbf{c}_i\|^2 \leq \|\mathbf{x}_j - \mathbf{c}_k\|^2, \text{ for each } k \neq i \\ 0 \text{ otherwise.} \end{cases} \tag{3.3}$$

Restated, *Xj*, belongs to group *i* if *Ci* is the closest center among all centers. Since a given data point can only be in a group, the membership matrix U has the following properties:

$$\sum_{i=1}^{c} u_{ij} = 1, \forall j = 1, \ldots, n$$

And

$$\sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij} = n.$$

On the other hand, if uij is fixed, then the optimal center *Ci* that minimize Equation (3.1) is the mean of all vectors in group i:

$$\mathbf{c}_i = \frac{1}{|G_i|} \sum_{k, \mathbf{x}_k \in G_i} \mathbf{x}_k \tag{3.4}$$

Where $|Gi|$ is the size of Gi, or $|Gi| = \sum_{j=1}^{n} Uij$.

For a batch-mode operation, the K-means algorithm is presented with a data set Xj, i=1,..., n; the algorithm determines the cluster centers *Ci* and the membership matrix U iteratively using the following steps:

Step 1: Initialize the cluster center *Ci*, i = 1…..c. This is typically achieved by randomly selecting c points from among all of the data points.

Step 2: Determine the membership matrix U by Equation (3.3).

Step 3: Compute the cost function according to Equation (3.1). Stop if either it is below a certain tolerance value or its improvement over previous iteration is below a certain threshold.

Step 4: Update the cluster centers according to Equation (3.4). Go to step 2.

The algorithm is inherently iterative, and no guarantee can be made that it will converge to an optimum solution. The performance of the K-means algorithm depends on the initial positions of the cluster centers, thereby making it advisable either to employ some front-end methods to find good initial cluster centers or to run the algorithm several times, each with a different set of initial cluster centers. Moreover, the preceding algorithm is only a representative one; it is also possible to initialize a random membership matrix first and then follow the iterative procedure.

The K-means algorithm can also be operated in the on-line mode, where the cluster centers and the corresponding groups are derived through time averaging. That is, for a given data point x, the algorithm finds the closest cluster center *Ci* and it is updated using the formula

$$\Delta C_i = \eta (X - C_i)$$

This on-line formula is essentially embedded in many learning rules of the unsupervised learning neural networks .

3.1.2 Fuzzy C-Means Clustering

Fuzzy C-means clustering (FCM), also known as fuzzy ISODATA, is a data clustering algorithm in which each data point belongs to a cluster to a degree specified by a membership grade. Bezdek proposed this algorithm in 1973 as an improvement over earlier hard C-means (HCM) clustering described in the previous section.

FCM partitions a collection of n vector *Xj* ,z = 1, ..., n into c fuzzy groups, and finds a cluster center in each group such that a cost function of dissimilarity measure is minimized. The major difference between FCM and HCM is that FCM employs fuzzy partitioning such that a given data point can belong to several groups with the degree of belongingness specified by membership grades between 0 and 1. To accommodate the introduction of fuzzy partitioning, the membership matrix U is allowed to have elements with values between 0 and 1. However, imposing normalization stipulates that the summation of degrees of belongingness for a data set always be equal to unity:

$$\sum_{i=1}^{c} u_{ij} = 1, \forall j = 1, \ldots, n. \qquad (3.5)$$

The cost function (or objective function) for FCM is then a generalization of Equation (3.1):

$$J(U, \mathbf{c}_1, \ldots, \mathbf{c}_c) = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c} \sum_{j}^{n} u_{ij}^m d_{ij}^2 \tag{3.6}$$

Where *Uij* is between 0 and 1; *Ci* is the cluster center of fuzzy group i; d*ij* = ||*Ci* —*Xj*|| is the Euclidean distance between iih cluster center and jth data point; and m € G [l, 0) is a weighting exponent.

The necessary conditions for Equation (3.6) to reach a minimum can be found by forming a new objective function J as follows:

$$\begin{aligned} \bar{J}(U, \mathbf{c}_1, \ldots, \mathbf{c}_c, \lambda_1, \ldots, \lambda_n) &= J(U, \mathbf{c}_1, \ldots, \mathbf{c}_c) + \sum_{j=1}^{n} \lambda_j \left(\sum_{i=1}^{c} u_{ij} - 1\right) \\ &= \sum_{i=1}^{c} \sum_{j}^{n} u_{ij}^m d_{ij}^2 + \sum_{j=1}^{n} \lambda_j \left(\sum_{i=1}^{c} u_{ij} - 1\right) \end{aligned} \tag{3.7}$$

where Xj, j = 1 to n, are the Lagrange multipliers for the n constraints in Equation (3.5). By differentiating J(U, C1,..., Cc, Ai,..., An) with respect to all its input arguments, the necessary conditions for Equation (3.6) to reach its minimum are

$$\mathbf{c}_i = \frac{\sum_{j=1}^{n} u_{ij}^m \mathbf{x}_j}{\sum_{j=1}^{n} u_{ij}^m}, \tag{3.8}$$

and

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{kj}}\right)^{2/(m-1)}} \tag{3.9}$$

The fuzzy C-means algorithm is simply an iterated procedure through the preceding two necessary conditions. In a batch-mode operation, FCM determines the cluster centers *Ci* and the membership matrix U using the following steps [1]:

Step 1: Initialize the membership matrix U with random values between 0 and 1such that the constraints in Equation (3.5) are satisfied.

Step 2: Calculate c fuzzy cluster centers *Ci*, i = 1,... , c, using Equation (3.8).

Step 3: Compute the cost function according to Equation (3.6). Stop if either itis below a certain tolerance value or its improvement over previous iteration is below a certain threshold.

Step 4: Compute a new U using Equation (3.9). Go to step 2.

The cluster centers can also be first initialized and then the iterative procedure carried out. No guarantee ensures that FCM converges to an optimum solution. The performance depends on the initial cluster centers, thereby allowing us either to use another fast algorithm to determine the initial cluster centers or to run FCM several times, each starting with a different set of initial cluster centers. Figure 3.1 presents a MATLAB demo of the fuzzy C-means clustering method in the Fuzzy Logic Toolbox. The data set, number of clusters, exponent weighting, and several stopping criteria can all be changed via the graphical user interface. Pushing the "Start" button allows one to observe how the cluster centers move toward the "right" positions. In particular, if the "Label Data" is marked, you will be able to see how each group evolves when the cluster centers move. After the clustering process stops, a cluster center can be selected, which will display the membership grades of all data points toward the selected cluster center.

Bezdek's monograph provides a detailed treatment of fuzzy C-means clustering, including its variants and convergence properties. Applications of fuzzy C-means include medical image segmentation and qualitative modeling.

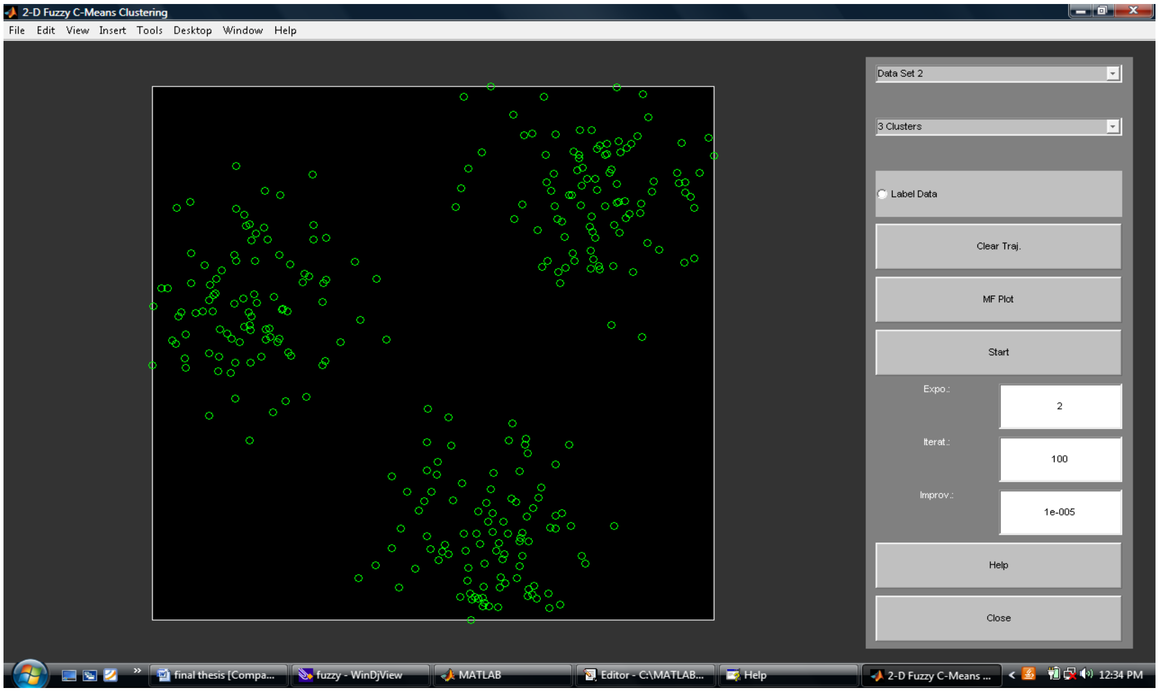

Figure 3.1Demo program for Fuzzy C- means Clustering algorithm (MATLAB Command: f cmdemo)

3.2 Summary

This chapter presents some of the most representative off-line clustering techniques frequently used in conjunction with radial basis function networks and fuzzy modeling: (hard) C-means clustering, fuzzy C-means clustering. These clustering techniques provide batchmode approaches to finding prototypes characterizing a data set; these prototypes are then used as the centers for radial basis functions in RBFNs or fuzzy rules in ANFIS. For data compression, these prototypes are used as a codebook in vector quantization [11].

CHAPTER 4

DATA COLLECTION AND REPRESENTATION

4.1 Data Collection

To provide traffic information there are several collaborative efforts from private company, government agencies to collect traffic information from major roads and report traffic condition to the public. They collect the traffic information through loops, camera, electronic toll tags, which are very effective but very expensive system. GPS data on a particular day provided by Dr. Rajeev Saraf, C.E.O., Lepton Software Export & Research Pvt. Ltd, GIS of Mumbai and the 'MapInfo' software which is required for work.

So we need an alternate way to collect traffic data at a lower cost with wider coverage to estimate the traffic situation.

Traffic Incident Detection Systems (IDS) is an area of Intelligent Transport System(ITS) use a variety of technologies to detect incidents so that bottlenecks created by incidents or hotspots can be cleared quickly. The success of all the Incident Detection System depends on the quality and range of real-time traffic data. The use of probe vehicle in IDS has become an attractive alternative to collect the data.

In this Thesis report, we use GPS technique to collect the data to identify the various hot-spots of the city. Using this technique, a GPS receiver is connected to a portable computer and collects the latitude and longitude information that enables tracking of the test vehicle. Also with longitude and latitude of the probe vehicle, we also collected the vehicle id, their speed and their status that is either the vehicle is running or stopped at particular time [12,14].

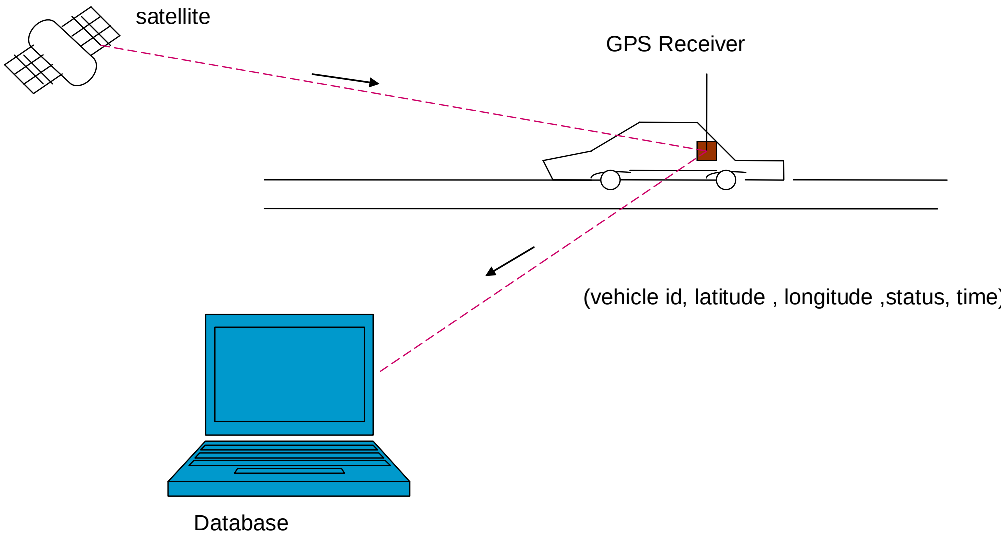

Figure 4.1: GPS enabled probe vehicle

For example: we have data of 23 March in the form…….

Vehicleid	stringdatetime	Vehstatus	speed	Lat	Lon
357023002922368	23-March-10	S	0	19.1797734	72.835045
357023002922368	23-March-10	S	0	19.1797533	72.834941
357023002922368	23-March-10	S	0	19.1795633	72.835255
357023002922368	23-March-10	S	0	19.1797033	72.835190
357023002922368	23-March-10	S	0	19.1798467	72.835135
357023002922368	23-March-10	S	0	19.1797667	72.835135

we have data of 24 March in the form…….

Vehicleid	stringdatetime	Vehstatus	speed	Lat	Lon
357023002922368	24-March-10	S	0	19.1797742	72.835145
357023002922368	24-March-10	S	0	19.1797122	72.834412
357023002922368	24-March-10	S	0	19.1795722	72.835422
357023002922368	24-March-10	S	0	19.1797001	72.835132
357023002922368	24-March-10	S	0	19.1798261	72.835213
357023002922368	24-March-10	S	0	19.1797777	72.835327

we have data of 25 March in the form…….

Vehicleid	stringdatetime	Vehstatus	speed	Lat	Lon
357023002922368	25-March-10	S	0	19.1797541	72.835121
357023002922368	25-March-10	S	0	19.1797234	72.834321
357023002922368	25-March-10	S	0	19.1795711	72.835212
357023002922368	25-March-10	S	0	19.1797435	72.835231
357023002922368	25-March-10	S	0	19.1798543	72.835143
357023002922368	25-March-10	S	0	19.1797762	72.835154

we have data of 26 March in the form…….

Vehicleid	stringdatetime	Vehstatus	speed	Lat	Lon
357023002922368	26-March-10	S	0	19.1797723	72.835012
357023002922368	26-March-10	S	0	19.1797543	72.834923
357023002922368	26-March-10	S	0	19.1795656	72.835221
357023002922368	26-March-10	S	0	19.1797032	72.835123
357023002922368	26-March-10	S	0	19.1798467	72.835167
357023002922368	26-March-10	S	0	19.1797689	72.835165

4.2 Mapping Of Data:

We have mapped the GPS data of different dates on to the road map of a city for which we need the traffic information and the identification of the 'hot-spots'.

For example … small dots on the city map are the GPS data on a particular day.

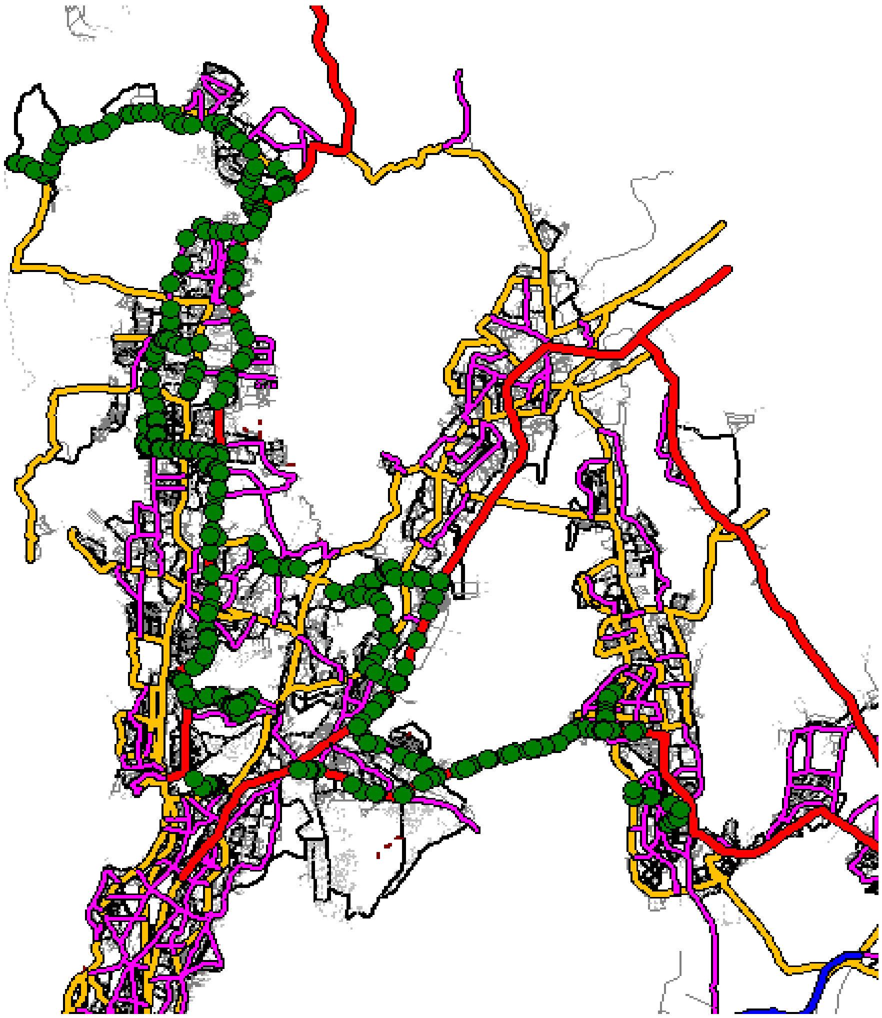

Figure 4.2: GPS data mapped on the road map of Mumbai

CHAPTER 5

ANALYSIS OF DATA

They collect the traffic information through loops, camera, electronic toll tags, which are very effective but very expensive system.

Identifying traffic pattern is quite difficult, identifying individual vehicle behavior is even more difficult, as it involves analyzing factors such as vehicle type, timing, speed, road type, location, conditions.

The traffic pattern is a cumulative behavior of vehicles, such as their number or the average speed of vehicles on certain road sections.

5.1 Proposed algorithm for identifying the Hot-spots

Here we present a method for detecting the location of traffic incidents or 'hot-spots', that causing the congestion in the normal traffic flow. The algorithm works in several steps. Below we explain the necessary steps of the algorithm.

Step 1-: To analyze the traffic behavior efficiently, segment the road (ie roads are divided into the grids) and after the segmentation assign a normal average speed to each road segment depends on the time, type of the road, and day. For example 500 meters segments for a motorway type on weekday's peek time with normal average speed between 45-60 Km/h under normal weather conditions.

Step 2-: Calculate the average speed of the vehicles in a certain direction in segment which has most no. of the GPS data.

$$\text{Current\_average\_speed} = (1/N) \sum \text{vehicle\_speed}$$

Where N is the no. of vehicle in that segment and summation runs over N.

Step 3-: If current _average_spped < Normal_average_speed, for any segment, then mark the segment for further analysis, and move to the next step.

Step 4-: Determine the current average speed of the road segments in front and behind of the marked segment, If the average speed in the 'front' segment is much higher than the 'marked' segment then a blockage within the slowest is likely. Since there are possibilities, that segments could include normal stoppage points such as traffic lights, junctions and roundabouts.

Step 5-: Compare the average speed of the vehicles in the neighboring segment with that of the marked segment.

Figure 5.1: Segment of road causing congestion

The average speed in segments in front and behind the marked segment is found in order to detect if there is an incident in the slowest marked segment or there is just a general congestion. If there is just a general congestion in the segment (as it consist of a place or is close to a place, where traffic is usually congested, for e.g. traffic lights, junction or level crossing) the average speed will be similar in adjoining segments. However, if there is an incident or hot spots causing a blockage in the marked segment then the average speed of the segments in front of that blocked/marked segment will be higher and with less number of vehicles.

5.1.1 Pseudo-Code of Proposed algorithm

Hot_Spot_Detector(City_Graph)

Step 1: Start

Step 2: Select a road from City_Graph

Step 3: Call SEGMENT(Road[m])

Step 4: For segment[i]= 1 to n step up by 1

Step 5: set cur_avg_speed = AVG_SPEED_CALC(segment[i])

Step 6: set normal_speed = avg_speed[segment[i]]

Step 7: if cur_avg_speed ≥ normal_speed

Step 8: Then goto step 4

Step 9: set front_avg_speed = AVG_SPEED_CALC(segment[i+1])

Step 10: if front_avg_speed ≤ avg_speed[segment[i+1]] + x

Step 11: then goto step 4

Step 12: if no_vehicle[segment[i+1]] < no_vehicle[segment[i]] - y

Step 13: then set hot_spot[segment[i]] = True

Step 14: End

SEGMENT(Road[M])

Step 1: Start

Step 2: divide the Road[m] into ‘n’ equal segments

Step 3: for segment[i] = 1 to n step up by 1

Step 4: initialize avg_speed[segment[i]]

Step 5: Return

AVG_SPEED_CALC(segmengt[x])

Step 1: Start

Step 2: set N= cur_no_vehicle[segment[x]] // collected from GPS data

Step 3: set cur_avg_speed = 0

Step 4: for each vehicle in segment[x] = 1 to N step up by 1

Step 5: set cur_avg_speed = cur_avg_speed + speed_vehicle[i]

Step 6: Return cur_avg_speed = cur_avg_speed/N

5.2 Results

- From the analysis of the GPS data of different days, we have identified some of the 'hot-spots', locations where traffic congestion has been taken place, on different days of the week.
- Also we have estimated the duration, in which these 'hot-spots' have impact on the traffic pattern, on different days of week.

Each identified 'hot-spots' are in the form of (location, duration, affected area).

Here is the list of different 'hot-spots' on different days as the result of above algorithm.

On 16th March-

TIME		LOCATION	DURATION	AREA
FROM	TO		(hrs.)	(m)
7:21:18 AM	8:53:51 AM	Samrudhhi Commercial Complex	1.32	700
11:35:04 AM	12:40:14 PM	Kripa Shankar Tower	1.05	650
1:10:15 AM	2:42:12 PM	Jagdamba complex	1.32	800
5:14:35 PM	7:30:12 PM	Samrudhhi Commercial Complex	2.16	1200

On 17th March –

TIME		LOCATION	DURATION	AREA
FROM	TO		(hrs.)	(m)
7:01:28 AM	8:15:32 AM	Samrudhhi Commercial Complex	1.14	900
9:15:24 AM	11:20:19 PM	Infinity Tower	2.05	1100
4:10:42 PM	5:00:45 PM	UTI Bank near Surya Hospital	.50	570
7:14:37 PM	9:10:38 PM	Samrudhhi Commercial Complex	1.56	800

On 18th March -

TIME		LOCATION	DURATION	AREA
FROM	TO		(hrs.)	(m)
6:40:18 AM	8:45:51 AM	Samrudhhi Commercial Complex	2.05	1200
10:30:13 AM	11:42:26 PM	Bank of India	1.12	600
12:23:31 PM	2:32:01 PM	Samrudhhi Commercial Complex	2.09	1300
4:06:57 PM	5:27:13 PM	Mumbai Mahanagar Palika	1.21	850

On 21st March –

TIME		LOCATION	DURATION	AREA
FROM	TO		(hrs.)	(m)
10:25:45 AM	11:09:53 AM	UTI Bank	.44	800
1:13:11 PM	2:15:44 PM	Mohras Tower	1.02	1100
4:06:57 PM	5:02:13 PM	Sankat Mochan Mandir Near Hotel Visava	.56	500
9:09:08 PM	11:10:47 PM	Samrudhhi Commercial Complex	2.01	1250

On 23rd March –

TIME		LOCATION	DURATION	AREA
FROM	TO		(hrs.)	(m)
7:24:15 AM	8:30:25 AM	Samrudhhi Commercial Complex	1.06	1000
8:45:42 AM	9:49:30 AM	Sankat Mochan Mandir Near Hotel Visava	1.04	550
11:10:52 AM	12:21:16 PM	Sai Nath Hotel	1.11	900
3:39:20 PM	4:58:53 PM	UTI Bank	1.19	1200
10:34:07 PM	11:37:38 PM	Gayakwad Bus Terminal	1.03	750

Impact of hot-Spots on Area:

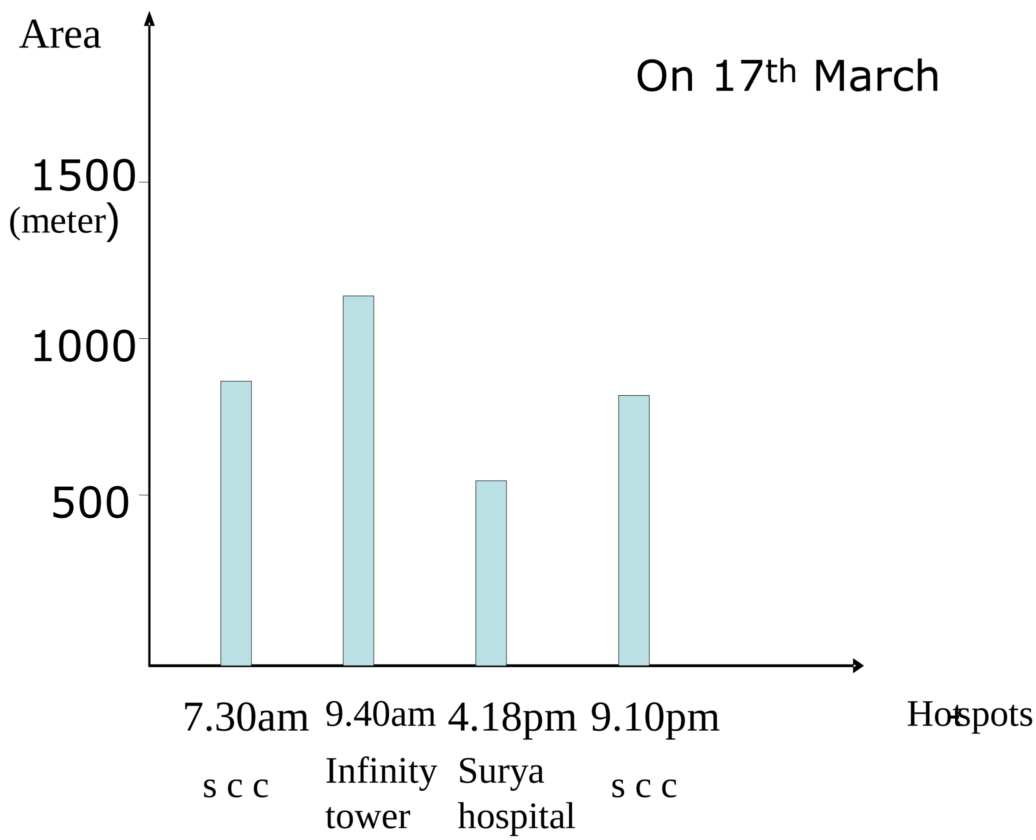

Figure 5.2: Impact of hot-spots on area on 17th March

Impact of hot- Spots on Time:

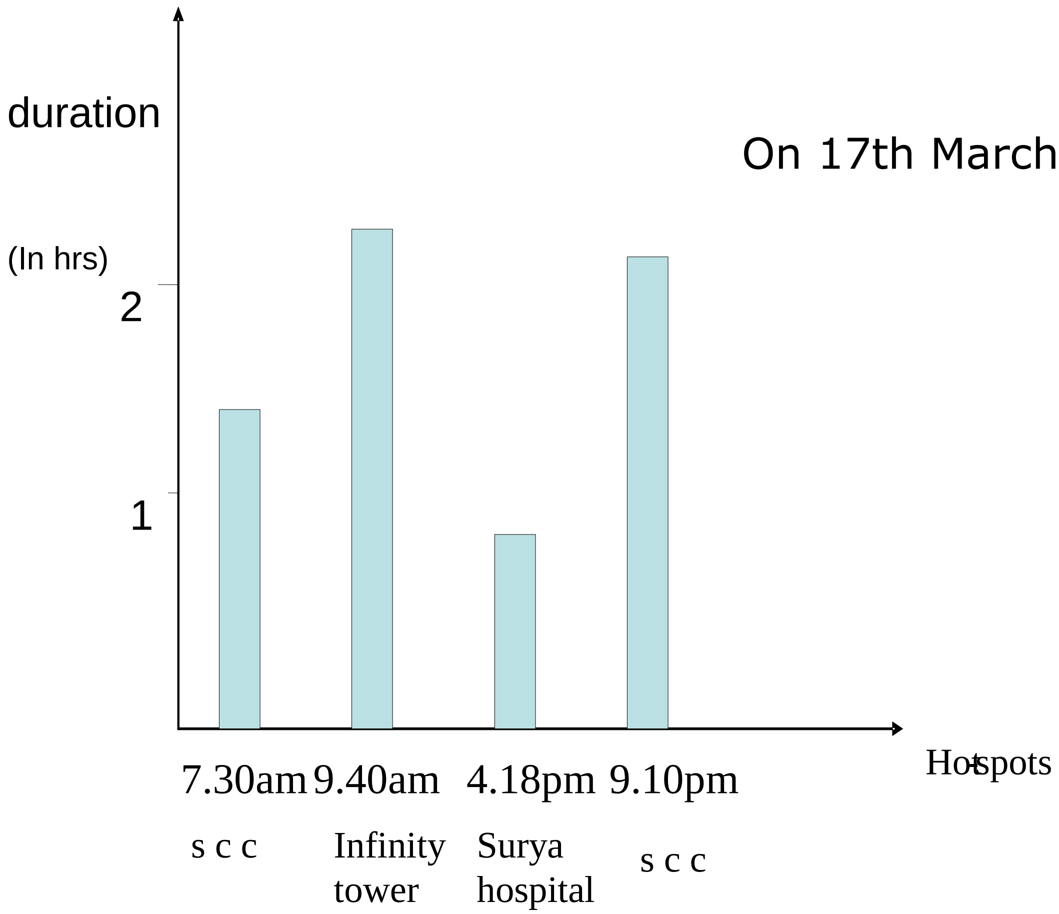

Figure 5.3: Impact of hot-spots on time on 17th March

CHAPTER 6

CLUSTERING ALGORITHM FOR DETECTION OF HOT- SPOTS

6.1 Introduction

Clustering is an optimization problem described by the objective function, which minimize distances between data points of one cluster and maximize distances between points belonging to two or more clusters. We use clustering algorithm for the detection of the 'hot-spots', where each cluster represents the group of GPS data points having latitude and longitude as their co-ordinate and very small distance between them. If there is congestion in traffic anywhere at any time, then GPS data points are very near to each other for some time interval, so we can apply the clustering algorithm for the detection of such congested area [11].

In hard clustering, data is divided into distinct cluster; where each data is belong to exactly one cluster. In fuzzy clustering, we associate a membership value to each data, which indicate the strength of association between that data element and a particular cluster. In this study fuzzy c-means algorithm is used for the detection of the 'hot-spots', which is most widely used in case of location problem[15,16].

6.2 C-means algorithm [11]

Fuzzy c-means clustering attempts to partition a finite collection of element into a collection of c-fuzzy cluster with respect to some criteria. Let x_i ε R^p, i=1, 2…..N, denote the data elements (represented as *n* real-valued columns vectors of dimension p). Let c_j ε R^p, j=1, 2…..C, represents the center of the cluster, with $2 \le C \le N$. Let U ε R^{CxN} denote the partition matrix comprised of fuzzy memberships.

The elements of U satisfy the following constraints:

(i) $0 \leq u_{ij} \leq 1$ (1)

(ii) $\sum_{j=1}^{C} u_{ij} = 1$ (2)

The fuzzy c-means clustering is based on the following optimization function, under the constraints of above equations:

$$\min_{u,c} \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^{m} \| x_i - c_j \|^2 , \qquad 1 \leq m \leq \infty \qquad (3)$$

Where u_{ij} is the degree of membership of x_i in the cluster j, and ||.|| is any norm expressing the similarity between any measured data and the center, here our norm is the distance between two points having co-ordinate as latitude and longitude, m is any real number greater than 1, and it controls the amount of "fuzziness". As m approaches 1, the fuzzy clusters become crisp clusters, where each data point belongs to only one cluster. A large value of m results in smaller membership value and as m approaches infinity, the clusters become completely fuzzy, and each point will belong to each cluster to the same degree (1/c) regardless of the data. The value of m is usually in the interval [1.5, 2.5].

The cluster centers c_j can be measured by:

$$c_j = \sum_{i=1}^{N} u_{ij}^{m} . x_i / \sum_{i=1}^{N} u_{ij}^{m} \qquad (4)$$

Fuzzy partitioning is carried out through an iterative optimization of the objective function, with the update of membership u_{ij} and the cluster centers c_j by:

$$u_{ij} = 1 / \sum_{k=1}^{C} (\| x_i - c_j \| / \| x_i - c_k \|)^{2/m-1} \qquad (5)$$

Iterative algorithm as follows.

Step 1- At first we initialize the partition matrix U = $\{u_{ij}\}$ comprised of fuzzy memberships, $U^{(0)}$.

Step 2- At k-th iteration: we calculate the center vectors $C^{(k)} = [\ c_j\ ]$ with $U^{(k)}$, by using the equation (4).

Step 3- At this step we will update partition matrix of fuzzy membership $U^{(k)}$ to $U^{(k+1)}$,by using the equation (5).

Step 4- At this step, we will check the stopping condition:

$$\text{Max } \{|\ U^{(k+1)} - U^{(k)}|\} < \varepsilon$$, where ε is a small number between 0 and 1.

If this condition is satisfied we will stop, otherwise we will go to step 2.

Once the stopping condition is satisfied, we have the cluster center in the form of (longitude, latitude) which represent the location of 'hot-spots'. We map these points on the map of the city to know about the area which are congested.

6.3 Measuring distance between two points on earth surface

The norm which we have used in the above algorithm is the distance between two points. Since earth is not purely spherical, it is a kind of ellipsoid. We have derived a formula for calculating geodesic distance between a pair of latitude/longitude points on the earth's surface, using an accurate WGS-84 ellipsoidal model of the earth, which is most accurate and widely used globally-applicable model for the earth.

a, b = major & minor semi axes of the ellipsoid

Where a = 6378137 m (±2 m) and b = 6356752.3142 m

f = flattening $(a - b) / a$

$\varphi 1, \varphi 2$ = geodetic latitude

L = difference in longitude

$U1 = \text{atan}((1-f).\tan\varphi 1)$, $U2 = \text{atan}((1-f).\tan\varphi 2)$

$\lambda = L$ (first approximation)

Iterate until change in λ is negligible

{

$\sin\sigma = \sqrt{[(\cos U2.\sin\lambda)^2 + (\cos U1.\sin U2 - \sin U1.\cos U2.\cos\lambda)^2]}$

$\cos\sigma = \sin U1.\sin U2 + \cos U1.\cos U2.\cos\lambda$

if $\sin\sigma = 0$ 'co-incident points', return;

$\sigma = \text{atan2}(\sin\sigma, \cos\sigma)$

$\sin\alpha = \cos U1.\cos U2.\sin\lambda / \sin\sigma$

$\cos^2\alpha = 1 - \sin^2\alpha$

$C = f/16.\cos^2\alpha. [4+f. (4-3.\cos^2\alpha)]$

$\lambda' = L + (1-C).f.\sin\alpha. (\sigma + C.\sin\sigma)$

}

$\cos 2\sigma m = \cos\sigma - 2.\sin U1.\sin U2/\cos^2\alpha$

$u^2 = \cos^2\alpha. (a^2 - b^2) / b^2$

$\Delta\sigma = b.\sin\sigma \{\cos 2\sigma m + b/4[\cos\sigma. (-1+2.\cos^2 2\sigma m) - b/6. \cos 2\sigma m. (-3+4\sin^2\sigma)]\}$

$s = b. (\sigma - \Delta\sigma)$

s is the required distance between two points.

Since trigonometric functions take arguments in radians, so latitude, longitude in degrees (either decimal or degrees/minutes/seconds) converted to radians, by using the formula

$$rad = \pi.deg/180.$$

Here we have used the trigonometric function atan2(), it takes two arguments, and computes the arc tangent of the ratio y/x. It is more flexible than atan(y/x), since it handles x=0, and it also returns values in all 4 quadrants $-\pi$ to $+\pi$. Also we have considered the case of co-incident points.

6.6 Results

Here are some results of the above algorithm.

1) Input: we have input the GPS data of 21st March.

Output: as the result of above algorithm, we find the following location as the ‘hot-spots’.

Latitude	Longitude
19.1026	72.9198
19.1396	72.8690
19.4060	72.8609
19.1788	72.8347

We map these points, on the road map of the Mumbai, to know the location of the ‘hot-spots’. Following figure shows the required location.

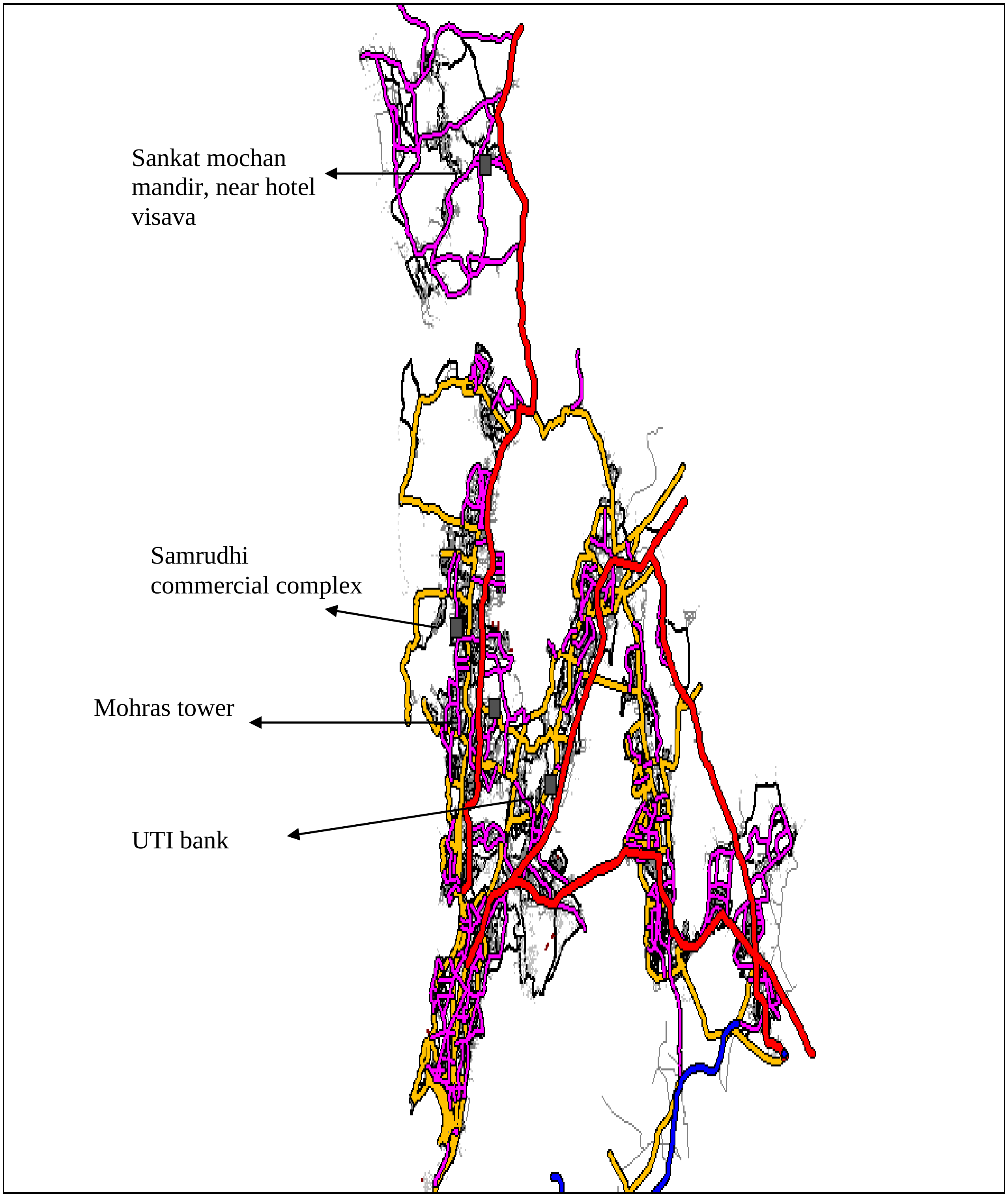

Fig: 6.1: Output the GPS data of 21st March

2) Input: we have input the GPS data of 23rd March.

Output: as the result of above algorithm, we find the following location as the 'hot-spots'.

Latitude	Longitude
19.3925	72.8298
19.3890	72.8265
19.2022	72.8428
19.2154	72.8385
19.1837	72.8197

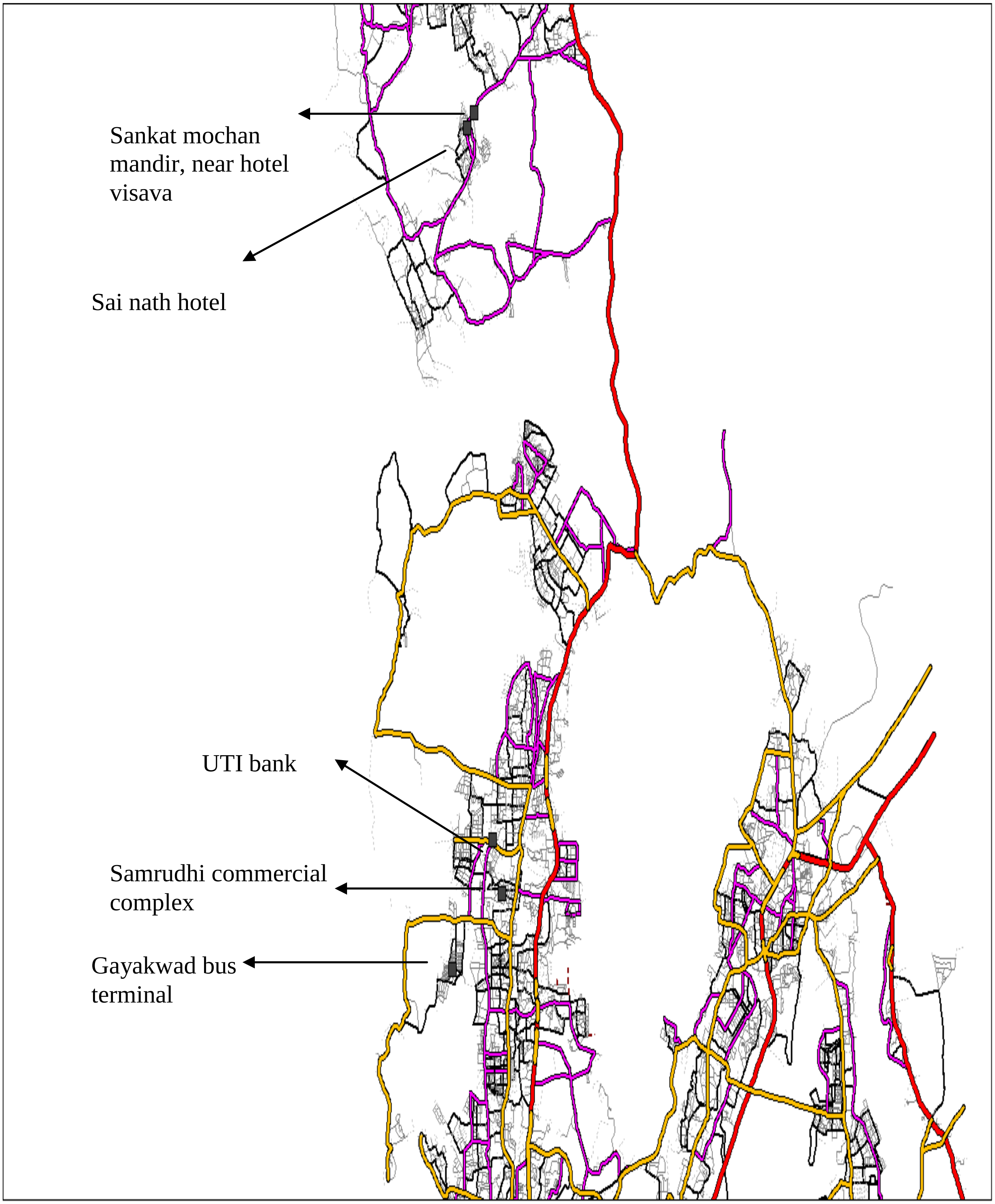

Fig: 6.2: Output the GPS data of 23st March

CHAPTER 7

CONCLUSIONS AND DISCUSSION

7.1 Concluding Remarks

This work focuses on the detection of the hot-spots (those areas of the city, where traffic is regularly congested and motorist tries to avoid these areas on their trip).Global Positioning System (GPS) data is used for this objective. Two algorithms are used for this purpose; one is based on the speed of the probe vehicle and other is the clustering algorithm. We have used Fuzzy c-means clustering for this purpose. Locations which we are getting as the result of both the algorithm are very much same.

The findings from this study are expected to help road users, while they are on their trip. Information about the 'hot-spots' can be provided to them by traffic management system, so they can choose alternative root if available, by this way we can reduce the severity of congestion.

7.2 Future Scope

Estimation of ‘hotspots’ is very important in modern day city planning because of the ever-increasing pressure of traffic on public travel. In particular, as many people have to travel/commute for their daily earnings identification of hotspots can save a lot of travel time. In particular we feel that the following points, further works need to be done:

1. One should use the techniques to identify ‘hotspots’ in different cities, and see the efficiency/correctness of our algorithms.
2. How can ‘hot-spots’ are utilized in performance monitoring, evaluation, planning, and management of road traffic system more efficiently.
3. Detection of the ‘Hot-Spots’, based on the clustering (Fuzzy c-means algorithm) algorithm.
4. To study how ‘hot-spots’ can be used in designing efficient of path finding algorithms in travel forecasting models, especially in calculating delays in important road junctions and at important time points.
5. To study how hot-spots’ can be used in designing efficient of path finding algorithms in travel of IMP/VIP probe Vehicles.

We expect that these studies will be taken up in future.

REFERENCES

[1] http:// www.gps.gov

[2] J.-S. Yang, "A Neural Network-Based Simultaneous Perturbation Stochastic Approximation Approach to Traffic Signal Timing Control," *WSEAS* Transactions on Systems, vol. 4, no. 3, pp. 197-205, 2005.

[3] Shoaib Kmaran, “A Multilevel traffic Incidents Detection Approach: Using real time GPS data”, Proceedings of the 2007 IEEE Intelligent vehicles Sympossium, Istanbul, Turkey, June 13-15, 2007.

[4] C. Basnayake, “Automated traffic incident detection with equipped probe vehicles,” Proceedings of the 17th International Technical Meeting of the Satellite Division of the Institute of Navigation, 741-750, 2004.

[5] Y. Li, and M. McDonald , “Motorway incident detection using probe vehicles,” Proceedings of the Institution of Civil Engineers, Transport, vol.158, pp.11-15, February 2005.

[6] Yuv-Horng Wen and Tsu- Tian Lee “Mining Fuzzy Rules for a Traffic Information System,”7th International Conference on Knowledge Based Intelligent Information and Engineering Systems, LNAI 2773, Oxford, pp. 237-243, 2005.

[7] K. Ozbay, and P. Kachroo, “Incident Management in Intelligent Transportation Systems,” London: Artech House Inc., pp.41-59, 1999.

[8] Niittymaki, J., and S. Kikuchi. “Application of Fuzzy Logic to the Control of a Pedestrian Crossing Signal”. In Transportation Research Record: Journal of the Transportation Research Board, No. 1651, TRB, National Research Council, Washington, D.C., 1998, pp. 30-38. Washington, D.C., 1998, pp. 30-38.

[9] Chiu, S. Adaptive Traffic Signal Control Using Fuzzy Logic. Proceedings of the IEEE Intelligent Vehicles Symposium, 1992, pp. 98-107.

[10] Niittymäki J., Mäenpää M. (2001). The role of fuzzy logic public transport priority in traffic signal control. Traffic Engineering and Control, International Journal of Traffic Management and Transportation Planning, January 2001, Hemming-Group Ltd. By permission.

[11] Jyh-Shing Roger Jang, Chuen-Tsai Sun & Eiji Mizutani Book “Neuro-Fuzzy and Soft Computing” by Prentice-Hall, Inc.Simon & Schuster/A Viacom Company Upper Saddle River, NJ 07458,1997,pp. 423-428.

[12] M.W. Sermons, and F.S. Koppelman, “Use of vehicle positioning data for arterial incident detection,” Transportation Research- Part C, vol.4, pp.87-96, 1996.

[13] J. Gallagher, “Travel time data collection using GPS,” National Traffic Data Acquisition Conference, Albuquerque, New Mexico, pp. 147-161, May 1996.

[14] Balke K., Dudek C. L. and Moutain C. E., “Using probe-measured travel times to detect major freeway incident in Houston, Texas,” Transportation Research Record 1554, 213–220, 1996.

[15] Hellinga B. and Knapp G., “Automatic vehicle identification technology-based freeway incident detection,” Transportation Research Record 1727, 213–220, 2000.

[16] R. Weil, J. Wootton, and A. Garcia-Ortiz, “Traffic incident detection: Sensors and algorithms,” Mathematical and Computer Modeling, vol.27, pp.257-291, May 1998.

PAPER PUBLISHED

- Deepak Garg, Jitesh P. Tripathi "Algorithm for Detection of Hot Spots of Traffic through Analysis of GPS Data" Annual Journal of IEEE DS BEACON, January, pp. 09-15, 2010.