Atmospheric Change on the Geographical Theme Finding Of Different Functions on Human Mobility

S. Ramana¹, S. Sabitha², R. Senthil Kumar³, T. Senthil Prakash⁴

¹Department of CSE, Shree Venkateshwara Hi-Tech Engineering College, Gobi
²Department of CSE, Shree Venkateshwara Hi-Tech Engineering College, Gobi
³Department of CSE, Shree Venkateshwara Hi-Tech Engineering College, Gobi
⁴Department of CSE, Shree Venkateshwara Hi-Tech Engineering College, Gobi

ABSTRACT- Over the last century, the planet has metaphorically contracted as transport has developed to meet the demands of the populous. Global participation in this expansion has been disproportionate as the driving force for transport demand is ultimately economic growth, which in itself results in an increased need for travel. The activities of the transport systems in most countries are sensitive to a range of weather extremes, including those related to precipitation, thunderstorms, temperature, winds, visibility and sea level. The impact of climate, climate variability and climate change, in particular the impact of these extremes on transport systems and adaptation measures are discussed. The foundation of climate services to assist informed decision-making for climate change adaptation and travelling time prediction, planning and designing, which require close collaboration among a wide range of Disciplines and the engagement of the users such as the transport systems’ communities by using the weather-traffic indices extracted have been validated to be surprisingly consistent with real world observations.

Keywords: Weather prediction, Traffic parameters count, Support vector machine, Neural networks, Factor analysis.

DATA MINING

Data mining or knowledge discovery in databases (KDD) is a collection of exploration techniques based on advanced analytical methods and tools for handling a large amount of information to extract the hidden information from the databases. These techniques can find novel patterns that may assist an enterprise in understanding the business better and in forecasting. Data mining is a collection of techniques for efficient automated discovery of previously unknown valid, novel, useful and understandable patterns in large databases. The patterns must be actionable so that they may be used in decision making process.

Almost all enterprise collects a variety of information electronically. In the recent years the cost of storage of data has reduced considerably. The enterprises have started to realize that the information accumulated over the years constitute important strategic asset and there is potential business intelligence hidden in the large volumes of data. This intelligence can be a secret weapon on which the success of business may depend on.

Databases today can range in size into the terabytes more than 1,000,000,000,000 bytes of data. Within these masses of data lies hidden information of strategic importance. But when there are so many trees, how do you draw meaningful conclusions about the forest? The newest answer is data mining, which is being used both to increase revenues and to reduce costs. The potential returns are enormous. Innovative organizations worldwide are already using data mining to locate and appeal to higher-value customers, to reconfigure their product offerings to increase sales, and to minimize losses due to error or fraud.

DATA MINING: THE PROCESS

Data mining is a process that uses a variety of data analysis tools to discover patterns and relationships in data that may be used to make valid predictions. The first and simplest analytical step in data mining is to describe the data summarize its statistical attributes (such as means and standard deviations), visually review it using charts and graphs, and look for potentially meaningful links among variables (such as values that often occur together).

As emphasized in the section on THE DATA MINING PROCESS, collecting, exploring and selecting the right data are critically important. But data description alone cannot provide an action plan. To overcome, a predictive model based on patterns, which are determined from known results is built, and then the model is tested on results outside the original sample. A good model should never be confused with reality just as a road map isn’t a perfect representation of the actual road, but it can be a useful guide to understanding the business. The final step is to empirically verify the model. For example, from a database of customers who have already responded to a particular offer, you’ve prospects are likeliest to respond to the same offer.
Fig 1: Mining data from a database.

DATA MINING & WAREHOUSING

Frequently, the data to be mined is first extracted from an enterprise data warehouse into a data mining database or data mart (Figure 1.2). There is some real benefit if the data is already part of a data warehouse. As it shall be seen later on, the problems of cleansing data for a data warehouse and for data mining are very similar. If the data has already been cleansed for a data warehouse, then it will not need further cleaning in order to be mined. Furthermore, many problems of data consolidation are already addressed and put in place maintenance procedures.

The data mining database may be a logical rather than a physical subset of the data warehouse, provided that the data warehouse DBMS can support the additional resource demands of data mining. If it cannot, then it will be better to do with a separate data mining database.

Fig 2: Data mining data mart extracted from a data warehouse.

A data warehouse is not a requirement for data mining. Setting up a large data warehouse that consolidates data from multiple sources, resolves data integrity problems, and loads the data into a query database can be an enormous task, sometimes taking years and costing millions of dollars. However, it is possible to mine data from one or more operational or transactional databases by simply extracting it into a read-only database (Figure 1.3). This new database functions as a type of data mart.

Fig 3: Data mining data mart extracted from operational databases.

URBAN COMPUTING

URBAN computing connects urban sensing, data management, data analytic and service providing into a recurrent process for an unobtrusive and continuous improvement of people’s lives, city operation systems and the environment. The aim is to solve a variety of emerging city problems, such as traffic congestion, energy consumption, and pollution, based on the data of traffic flow, human mobility, and geographical data, etc. In particular, many works have been done to investigate the impact of inclement weather to traffic. For example, a heavy rain may slow down the traffic and cause congestions due to low visibility and high demand of vehicles; the decreasing temperature in very cold days will freeze the roads and influence the transport performance, etc.

The general relevance of the impact of weather change to transport in US. In July 21st, 2012, Beijing faces its largest rainstorm since 1951, with an average rainfall of 164 millimeters. According to the news report, there are 77 people died in this catastrophic natural disaster. The transport of Beijing suffered from various contingencies due to the serious flood. During that time, a variety of photos titled “see the sea in Beijing” widespread on the Internet. This disaster not only shows the serious problems of the urban transport system of Beijing, but also inspires our research interest: how can we identify those regions being highly influenced by weather change on transport? The early works often focus on the correlation of weather and traffic in some particular roads where devices have been deployed to continuously collect traffic data.

By analyzing the traffic change in different weather conditions, the traffic prediction can be better preformed considering the weather forecast. However, the weather-traffic correlation covering most roads throughout a city (known as regional weather-traffic sensitivity index or for simplicity weather-traffic index) is still an open problem vain in spite of the practical value in our daily life. One essential reason is the lacking of effective traffic monitoring system in city-wide scale. Another open problem is how to disclose the key factors behind the weather-traffic index, to explain the reason why some regions in a city are more vulnerable to inclement weather and others are not. These factors are the regional features including the density of roads, the number of road intersections, the number of POIs (points of interest), the traffic volume, the average age of the household, the density of buildings and more in the surrounding regions.

WEATHER-TRAFFIC INDEX
The weather-traffic index throughout a city and the knowledge of key factors behind the correlation provides effective support to help government agents to understand the functional character of districts throughout a city to improve traffic performance and to learn the key factors in urban planning, etc. For example, if the traffic of a region is highly affected by heavy rains, and the key factors include the sewer system, then it is important for the government to examine and improve the sewer system of the region in first priority.

To enable weather-traffic index throughout a city and factor analysis, the effective traffic monitoring in city-wide scale is a must. Nowadays, with the widely commercial use of taxi tracking system, the most feasible means probably is to extract traffic information from numerous taxis driving on roads due to its availability, wide-coverage and low-cost. A taxi tracking system combines the use of automatic vehicle location in individual vehicles with software that collects these fleet data. Typically, taxi data continuously record the information including location, speed, occupancy status, and orientation of the taxis. The traffic parameters (e.g., traffic speed) extracted from taxi data are practically sparse since the number of taxis in a city is typically limited. Therefore, we partition the city by Voronoi diagram where the seeds are the road intersections.

Compared to the region-oriented city map partition approach such as equal-sized rectangles where the roads in some cells are highly dense and in others are highly sparse, the advantage of our road-intersection-oriented partition makes sure every cell include at least one road intersection and a number of roads connected to this intersection. Given a period of time, the average parameters of driving taxis in each Voronoi cell (or called cell for simplicity in the rest of this work) is extracted as the average traffic parameter of the cell. In addition to traffic data, weather data and complicated regional features in the same period of time are also required to perform the study.

WEATHER IMPACTS ON SURFACE TRANSPORTATION

Travellers advisories, issued by the National Weather Service when transportation-sensitive conditions may occur, necessarily vary by geographical area. A trace of snowfall would trigger a traveller’s advisory in the southern portions of the USA, but several centimetres of snow would be the threshold for an advisory in the northern regions, where drivers are accustomed to snowy conditions.

Since about one-quarter of transport delays and crashes in the USA are due to adverse weather, the need for region-specific types of surface weather information is reassessed regularly. Identifying the transportation-sensitive weather parameters was aided by Weather Information for Surface Transportation—National Needs Assessment Report, henceforth called WIST (Office of the Federal Coordinator for Meteorological Services and Supporting Research (OFCM), 2002).

Weather parameters, arranged by category of weather advisories, are grouped in the table on the previous page, along with their impacts. The WIST report established the national needs and requirements for weather information associated with decision-making for surface transportation operations in six different sectors: road; long-haul railway; marine transportation; rural and urban transit; pipeline systems; and airport ground operations. Here, we will take a look at railway, roadway and marine transport and the impacts of the two dominant weather factors: extreme temperature and extreme precipitation.

CLIMATE-CHANGE IMPACTS ON SURFACE TRANSPORTATION

Climate change will affect the efficiency, safety and reliability of existing transportation infrastructure and the design of new infrastructure. Retrofits and new designs will be expensive to implement, so sound climate data are essential to good decision-making. The climate projections used in the NRC study were based on both observational data and model output, using several different techniques. While a full description can be pursued through the references, the global climate model output used were those that contributed to the Fourth Assessment Report (AR4) of the Intergovernmental Panel on Climate Change (IPCC). For this study, three greenhouse-gas emissions scenarios were analysed and compared to model simulations of the recent past.

It does not reproduce the full list of parameters and impacts on surface land and water transportation. They are available in the annex to the NRC report. Instead, a simple list of the key phenomena likely to undergo change with global warming affecting transport is provided in the box overleaf. The primary way that climate change can impact transportation is through changes in extreme weather conditions. As noted, there are some positive impacts. On balance, however, because systems are adapted to their historical range of extremes, the majority of the impacts of events outside this range are expected to be negative.

In all transport sectors, weather extremes affect the efficiency of operations and safety and integrity of the infrastructure. Previous sections described weather-related transport impacts from extreme high temperature and extreme precipitation events. As the following discussion indicates, these events are expected, with high confidence, to occur more frequently in the future because of climate change. Additionally, some of the most significant regional transportation infrastructure vulnerabilities to climate change will occur in coastal areas.

TEMPERATURE PROJECTIONS

The climate model output provides projections for changes in mean temperature. But how do these changes in mean temperatures predict extreme temperatures that may
result from climate change? The programme of NOAA’s National Climatic Data Canter (NCDC) Probabilities of Temperature Extremes in the USA (NOAA, 1999 (CD ROM)) was used to estimate the probability that a threshold temperature will occur for one or more consecutive days and/or the probability that a threshold temperature will be exceeded for any number of days for a station in a given month or season, based on statistics from the observed climate combined with model projections.

**PROBLEM DEFINITION**

**Weather Prediction Problem Characteristics:** Weather prediction is a complicated procedure that includes multiple specialized fields of expertise. “There are only two methods to predict weather: the empirical approach and the dynamical approach”. Lorenz thus separated weather forecasting methodologies into two main branches in terms of numerical modelling and scientific processing (AI) of meteorological data.

**Forecast Types:** Weather forecast expert systems can be categorized as terminal aerodrome forecasts (TAFs), public forecasts and marine forecasts in terms of field of interest and scale. In addition to these, many specialized forecasts are produced for specific purposes (e.g. weather pollution forecasting, agricultural predictions) TAFs are required to be most precise both in terms of measurable weather conditions and in terms of timing.

**Importance of TAF:** Airports organize air traffic mainly based on local weather condition that has agile parameters that may change in a short time. These parameters such as horizontal visibility and vertical visibility (ceiling), fogginess, precipitation level etc. can be very serious threat for flight safety and cost. When ceiling and visibility at a busy airport are low, in order to maximize safety, the rate of planes landing is reduced. When ceiling and visibility at a destination airport are forecast to be low at a flight’s scheduled arrival time, its departure may be delayed in order to minimize traffic congestion and related costs.

Effects of changes in the cryosphere have been documented in relation to virtually all cryospheric components, with robust evidence that they are, in general, a response to the reduction of snow and ice masses due to enhanced warming.

**MOUNTAIN GLACIERS AND ICE CAPS, ICE SHEETS AND ICE SHELVES**

Effects of changes in mountain glaciers and ice caps have been documented in runoff (Kaser et al., 2003; Box et al., 2006), changing hazard conditions (Haebeli and Burn, 2002) and ocean freshening (Bindoff et al., 2007). There is also emerging evidence of present crustal uplift in response to recent glacier melting in Alaska (Larsen et al., 2005). The enhanced melting, as well as the increased length of the melt season of glaciers, leads at first to increased river runoff and discharge peaks, while in the longer time-frame (decadal to century scale), glacier runoff is expected to decrease (Jansson et al., 2003). Evidence for increased runoff in recent decades due to enhanced glacier melt has already been detected in the tropical Andes and in the Alps.

The formation of lakes is occurring as glaciers retreat from prominent Little Ice Age (LIA) moraines in several steep mountain ranges, including the Himalayas (see Box 5.4), the Andes and the Alps. Thawing of buried ice also threatens to destabilise the Little Ice Age moraines. These lakes thus have a high potential for glacial lake outburst floods (GLOFs). Governmental institutions in the respective countries have undertaken extensive safety work, and several of the lakes are now either solidly dammed or drained; but continued vigilance is needed, since many tens of potentially dangerous glacial lakes still exist in the Himalayas (Yamada, 1998) and the Andes (Ames, 1998), together with several more in other mountain ranges of the world.

Glacier retreat causes striking changes in the landscape, which has affected living conditions and local tourism in many mountain regions around the world (Watson and Haebeli, 2004; Mölg et al., 2005). Figure 5.10 shows the effects of the retreat of the Chacaltaya Glacier on the local landscape and skiing industry. Warming produces an enhanced spring–summer melting of glaciers, particularly in areas of ablation, with a corresponding loss of seasonal snow cover that results in increased exposure of surface crevasses, which can in turn affect, for example, snow runway operations, as has been reported in the Antarctic Peninsula (Rivera et al., 2005).

**SNOW COVER AND FROZEN GROUND**

Due to less extended snow cover both in space and time, spring peak river flows have been occurring 1–2 weeks earlier during the last 65 years in North America and northern Eurasia. There is also evidence for an increase in winter base flow in northern Eurasia and North America, as well as a measured trend towards less snow at low altitudes, which is affecting skiing areas.

Reductions in the extent of seasonally frozen ground and permafrost, and an increase in active-layer thickness, have resulted in: the disappearance of lakes due to draining within the permafrost, as detected in Alaska (Yoshikawa and Hinzman, 2003) and Siberia (see Figure 5.12) (Smith et al., 2005); a decrease in potential travel days of vehicles over frozen roads in Alaska; increased coastal erosion in the Arctic (e.g., Beaulieu and Allard, 2003)

**HYDROLOGY AND WATER RESOURCES**

Changes in surface and groundwater systems Since the TAR there have been many studies related to trends in river flows during the 20th century at scales ranging from catchment to global. Some of these studies have detected significant trends in some indicators of river flow, and some have demonstrated statistically significant links with trends in temperature or precipitation; but no globally homogeneous trend has been reported. Many studies, however, have found no trends, or have been unable to separate the effects of variations in temperature and precipitation from the effects of human interventions in the catchment, such as land-use
change and reservoir construction. Variation in river flows from year to year is also very strongly influenced in some regions by large-scale atmospheric circulation patterns associated with ENSO, NAO and other variability systems that operate at within-decadal and multi-decadal time-scales.

At the global scale, there is evidence of a broadly coherent pattern of change in annual runoff, with some regions experiencing an increase (Tao et al., 2003a, b, for China; Hyvärinen, 2003, for Finland; Walter et al., 2004, for the coterminous USA), particularly at higher latitudes, and others a decrease, for example in parts of West Africa, southern Europe and southern Latin America (Milly et al., 2005). Labat et al. (2004) claimed a 4% increase in global total runoff per 1°C rise in temperature during the 20th century, with regional variation around this trend, but this has been challenged due to the effects of non-climatic drivers on runoff and bias due to the small number of data points (Legates et al., 2005). Gedney et al. (2006) gave the first tentative evidence that CO2 forcing leads to increases in runoff due to the effects of elevated CO2 concentrations on plant physiology, although other evidence for such a relationship is difficult to find. The methodology used to search for trends can also influence results, since omitting the effects of cross-correlation between river catchments can lead to an overestimation of the number of catchments showing significant trends (Douglas et al., 2000).

Groundwater flow in shallow aquifers is part of the hydrological cycle and is affected by climate variability and change through recharge processes (Chen et al., 2002), as well as by human interventions in many locations (Petheram et al., 2001). [WGII 1.3.2.1] Groundwater levels of many aquifers around the world show a decreasing trend over the last few decades [WGII 3.2, 10.4.2], but this is generally due to groundwater pumping surpassing groundwater recharge rates, and not to a climate related decrease in groundwater recharge. There may be regions, such as south-western Australia, where increased groundwater withdrawals have been caused not only by increased water demand but also because of a climate-related decrease in recharge from surface water supplies (Government of Western Australia, 2003). In the upper carbonate aquifer near Winnipeg, Canada, shallow well hydrographs show no obvious trends, but exhibit variations of 3–4 years correlated with changes in annual temperature and precipitation (Ferguson and George, 2003). Owing to a lack of data and the very slow reaction of groundwater systems to changing recharge conditions, climate related changes in groundwater recharge have not been observed.

At present, no globally consistent trend in lake levels has been found. While some lake levels have risen in Mongolia and China (Xinjiang) in response to increased snow- and ice melt, other lake levels in China (Qinghai), Australia, Africa (Zimbabwe, Zambia and Malawi), North America (North Dakota) and Europe (central Italy) have declined due to the combined effects of drought, warming and human activities. Within permafrost areas in the Arctic, recent warming has resulted in the temporary formation of lakes due to the onset of melting, which then drain rapidly due to permafrost degradation (e.g., Smith et al., 2005). A similar effect has been reported for a lake formed over an Arctic ice shelf (i.e., an epishelf lake12), which disappeared when the ice shelf collapsed (Mueller et al., 2003). Permafrost and epishelf lakes are treated in detail by Le Treut et al. (2007).

**WATER QUALITY**

A climate-related warming of lakes and rivers has been observed over recent decades. As a result, freshwater ecosystems have shown changes in species composition, organism abundance, productivity and phenological shifts (including earlier fish migration). Also due to warming, many lakes have exhibited prolonged stratification with decreases in surface layer nutrient concentration and prolonged depletion of oxygen in deeper layers.

Due to strong anthropogenic impacts not related to climate change, there is no evidence for consistent climate-related trends in other water quality parameters (e.g., salinity, pathogens or nutrients) in lakes, rivers and groundwater. Thermal structure of lakes Higher water temperatures have been reported in lakes in response to warmer conditions. Heavy precipitation events has increased over most areas during the late 20th century, and that it is more likely than not that there has been a human contribution to this trend.

**FLOODS**

A variety of climatic and non-climatic processes influence flood processes, resulting in river floods, flash floods, urban floods, sewer floods, glacial lake outburst floods and coastal floods. These flood-producing processes include intense and/or long-lasting precipitation, snowmelt, dam break, reduced conveyance due to ice jams or landslides, or by storm. Floods depend on precipitation intensity, volume, timing, phase (rain or snow), antecedent conditions of rivers and their drainage basins (e.g., presence of snow and ice, soil character and status (frozen or not, saturated or unsaturated), wetness, rate and timing of snow/ice melt, urbanisation, existence of dykes, dams and reservoirs). Human encroachment into flood plains and lack of flood response plans increase the damage potential.

The observed increase in precipitation intensity and other observed climate changes, e.g., an increase in westerly weather patterns during winter over Europe, leading to very rainy low-pressure systems that often trigger floods (Kron and Berz, 2007), indicate that climate change might already have had an impact on the intensity and frequency of floods.

The Working Group I AR4 Summary for Policymakers concluded that it is likely that the frequency of heavy precipitation events has increased over most areas during the late 20th century, and that it is more likely than not that there has been a human contribution to this trend. Globally, the number of great inland flood catastrophes during the last 10 years (1996–2005) is twice as large, per decade, as between 1950 and 1980, while related economic
losses have increased by a factor of five (Kron and Berz, 2007).

Dominant drivers of the upward trend of flood damage are socio-economic factors such as economic growth, increases in population and in the wealth concentrated in vulnerable areas, and land-use change. Floods have been the most reported natural disaster events in many regions, affecting 140 million people per year on average (WDR, 2003, 2004). In Bangladesh, during the 1998 flood, about 70% of the country’s area was inundated (compared to an average value of 20–25%) (Mirza, 2003; Clarke and King, 2004).

**TRAFFIC PARAMETERS COUNTING OF VEHICLES**

Traffic Parameters Counting of vehicles is the most fundamental parameter in traffic, and most of the studies that are done in traffic engineering use this count to quantify traffic volume, flow and capacity of a section of a lane. Another important parameter is the occupancy of loop detectors, which is the percent of time vehicles are present on the detector. Traffic speeds measured at a single location are highly variable and not widely used. Speeds can be averaged over time (time mean speed) or space (space mean speed), but space mean speed is more meaningful in traffic analysis. Another important parameter is the travel time and commonly estimated from the flow and occupancy data. This section describes the basic traffic parameters.

**Traffic Volume** Traffic volume is defined as the number of vehicles passing through a point of a lane or a freeway in a given time interval (Transportation Engineers 1999). Traffic volume is one of the measures that estimate the amount of traffic flow at a given point. The volume of the traffic is often measured on annual, monthly, daily, hourly or sub minute intervals.

**Traffic Flow Rate** Traffic flow is the estimation of the number of vehicles that will pass through a point or a section of the freeway in one hour based on the count of the vehicles in the given interval. Flow rate is equivalent to hourly rate at which vehicles pass over a given point of a lane or a roadway at a given time instance (Roess et al 1997).

**Volume differs from flow rate** in that volume is the actual count of the vehicles that have passed through the point or the section of the freeway during a given time interval whereas traffic flow represents the number of vehicles that could pass in one hour at a time instance. Flow is a concept borrowed from the hydrodynamic theory and frequently used to model traffic dynamics.

**Occupancy** is the percentage of time vehicles are present at a given point (traffic detector) during the given time interval. For example, 10% of occupancy in a 30 second interval means that vehicles were present on the detector for 3 seconds during the 30 second interval.

**Traffic Density** Density is defined as the number of vehicles per mile for a section of a lane or a roadway for the given time interval. Density is computed by, density \( k \) flow \( q \) speed \( u \) where denotes flow in vehicles/hr, and denotes space mean speed in mph. If speed is unknown, it is often estimated where is the occupancy observed and \( g \) is a constant that quantifies an average vehicle length.

**Traffic Speed** Traffic speed for a given time interval is the average speed of every vehicle travelled in a section of the lane or the roadway. It is a space mean speed and computed as the ratio of traffic flow with the corresponding density for the given time interval. The speed that is estimated using this formula is often not accurate as the computation of density from occupancy uses inaccurate average vehicle length \( g \) (Sun, Ritchie, 1999). Accuracy of vehicle length measurements could increase the speed computation but such measurements are rarely available.

**URBAN TRANSPORTATION AND LAND USE**

The building of railroads and highways has influenced urban development, and conversely the growth of urban areas has influenced the development of road, air, and rail networks that facilitate travel within and across urban areas. As urban areas become vulnerable to climate change, addressing transportation issues in adaption and mitigation involves addressing the interactions between the sector and land use in cities. Urban land use planning is based on functional designation of land for different human purposes including economic and leisure activities. Categories of land use typically include residential, commercial, industrial, recreational, natural protection, institutions, parking, vacant land, and transportation or utilities “the distribution of infrastructure in the transport system creates opportunities for spatial interaction.” This “can be measured as accessibility”. Further, “the distribution of accessibility in space co-determines location decisions” and results in changes in land use. And a combination of energy intensity as well as demand for transportation systems determines the degree of carbon emissions.

Lower levels of fuel efficiency and increase in demand for transportation result in greater greenhouse gas emissions. Individual land uses within the urban fabric, including transport infrastructure, in great measure determine the urban form Urban form can be measured through various density measures – for dimensions such as population or economic activity – and the spatial location of activities and households. Urban population density is closely related to vehicle miles travelled (VMT). VMT is a key indicator for greenhouse gas emissions, and conversely, the greenhouse gas emissions from transport vehicles are related to VMT. The relationship is far more complex and other measures including the price of fuel, employment levels, trip origins and destinations and the size of cities have much influence on vehicle miles travelled.

**VARIATION IN WEATHER PARAMETERS**

Variability is considered in the following weather parameters:

1. Wind speed = Arithmetic mean of the 9am and 3pm wind speeds in kph
2. Daily Precipitation
3. Hours of sunshine over the day
4. Humidity
Apparent temperature in degree Celsius, Here, the variable „Apparent temperature” is considered as people’s perception is different to the measured air temperature. Apparent temperature is calculated as:

\[ \text{ATEMP} = T + 0.33 \times 6.105 \times (17.27 T / 237.7 + T) \times 0.01 H - 0.7W - 4.0 \]

where,
T = Air temperature in degrees Celsius. “T” is the arithmetic mean of the maximum, 9am and 3pm daily temperatures.
H = Relative humidity (%)
W = Wind speed (m/s), Here, it is the arithmetic mean of the 9am and 3pm wind speeds.

**CLIMATE HAZARDS TO URBAN TRANSPORTATION SYSTEMS**

The climate hazards that may pose the greatest challenges to urban transportation systems are different for different kinds of physical hazards cities in different geographical environments, and for different modes of transportation. Physical hazards from global climate change that affect urban transportation systems can be expressed in terms of changes in average values and often higher variability of temperature, precipitation, storm frequency and severity, coastal storm inundations especially in conjunction with sea level rise, and other climate processes. The vulnerability of transportation assets to these physical hazards, stemming from the existing spatial and economic organization of the city and its transport system, depends on a range of factors, including:

1. design and spatial layout of transport infrastructure – flood hazards tend to be worse for underground infrastructure; storms affect land, marine, and air travel differently;
2. basic urban form – high-density settlements with mass transit systems, or mostly motorized transportation with low densities;
3. availability of resources to keep transportation systems functioning in both disaster and non-disaster conditions.

**FORECAST MODELS ARIMA**

Models Search space definition for ARIMA model development was guided by the preliminary order estimation described in the section on preliminary ARIMA estimation. It is tested and verified the preliminary assessment that a pure non-seasonal moving average model was not appropriate. Selection criteria for pure moving average models do not minimize until an order of q = 10 and the best fits were measurably worse than the pure autoregressive and mixed non-seasonal models. Preliminary model fitting clearly verified the appropriateness of a seasonal MA model component.

Seasonal AR parameters degraded model fit in all cases. Also, including MA parameters beyond the first order made a significant improvement to model fit in none of the cases tried, and the estimated higher-order parameters were statistically insignificant in all cases. SPSS calculates the standard error, log likelihood, Akaike information criterion (AIC), and Swartz Bayesian criterion (SBC) statistic for each estimated ARIMA model. Standard error is given for information but is not specifically a selection criterion because it measures only fit of the model to the development set. The AIC and SBC statistics, on the other hand, are measures designed to predict model fit on future realizations of the modelled process. We emphasized the SBC statistic because studies have indicated that AIC tends to overestimate .

The SBC values close to the minimum to be equivalent, and in these cases we favoured a simpler model. It is suggested in the literature that models with SBC or AIC statistics within some constant c (c = 2 as a typical value) of the minimum SBC or AIC should be considered competitive.

**ADAPTATION**

Planners and other city officials must take the changing climate into consideration as they plan for the future. As mentioned previously, there are two approaches to planning for climate change. The first is mitigation, which involves decreasing the amount of greenhouse gases that we produce. Climate change mitigation is important, but its impact will not be realized until far in the future. The second approach, adaptation, has more short-term effects. Adaptation to climate change, as defined by the Intergovernmental Panel on Climate Change (IPCC) in 2001, is an adjustment in ecological, social or economic systems in response to observed or expected changes in climatic stimuli and their effects and impacts in order to alleviate adverse impacts of change. The IPCC wrote that adaptation is reactive or anticipatory, but Adger et al. argued that is instead both reactive and anticipatory as it is often catalyzed by events such as extreme weather and focuses on mitigating the potential impacts of those events in the future. Climate change adaptation for transportation infrastructure can take many different forms. One easy way that transportation agencies can adapt their infrastructure is via its normal lifespan. Because transportation infrastructure has a finite service lifespan, it can be cost-effective to simply replace or adapt infrastructure at the end of that life. The cost of upgrading or replacing can be more prohibitive. Cost of adaptation measures is often a barrier to implementing them. Short-term adaptation actions can be relatively inexpensive, but long-term adaptation strategy can involve many cost intensive objectives, which is an obvious deterrent to long-term adaptation planning. It is important for city officials to keep in mind the benefits of adaptation planning compared to the costs. Another measure that prepares an organization for climate change adaptation is to complete an asset inventory.

The Transportation Research Board recommends conducting an inventory because it is a low-cost way to start planning for adaptation. An inventory serves multiple purposes. It assesses what infrastructure is the most vulnerable to changing weather, but also creates a record of where each infrastructure component is in its lifespan and the timeline for replacing or repairing it. Overall, an inventory helps an agency to prioritize different parts of its system.
DISADVANTAGE OF EXISTING SYSTEM
1. To enhance the resilience of, both transport and flood risk management infrastructures are difficult.
2. How to identify regional weather-traffic sensitivity index throughout a city, that indicates the degree to which the region traffic in a city is impacted by weather changes.
3. Among complex regional features, such as road structure and population density. How to dissect the most influential regional features that drive the urban region traffic to be more vulnerable to weather changes.

WEATHER-TRAFFIC INDEX ESTABLISHMENT
The weather data and traffic data from data preparation component is the input of weather-traffic index establishment component. The intuition that the traffic is influenced by weather. It illustrates the average speeds in different cells in Shanghai at the same time slots in two different weather conditions: cloudy and rainy. It is clear that the average speed in rainy days is generally lower than that in cloudy days. At the same time, it also demonstrates that the average speeds in some cells are unchanged in the rainy days and in cloudy days. Therefore, weather-traffic index is necessary to indicate the impact of weather to traffic in different cells.

CORRELATION DETECTION
In a cell, for detecting the correlation between the traffic speed, denoted as Ft, and weather, denoted as Fw, a simple method is to train a classifier which infers directly from Fw to Ft.
1. The input is the weather represented as a feature vector and the output is one of the seven speed classes.
2. The trained classifier is tested. If the inference accuracy is high, it means the correlation between the traffic and weather is high in this cell;
3. The correlation is low.
4. This method is commonly used in statistics to measure the correlation between two random variables critical weakness of this method in correlation detection between traffic and weather.
5. This is because there are many other reasons which impact traffic.

Traffic prediction is a well studied problem. Since early 1980s, univariate time series models, mainly Box-Jenkins Auto Regressive Integrated Moving Average (ARIMA) and HoltWinters Exponential Smoothing (ES) , have been widely used in traffic prediction. In the last decade, neural network models have also been used in forecasting travel time. In spatial-temporal characteristics of traffic events are considered in training traffic prediction models. In authors use AQ21, a natural induction system that learns and applies attribution rules, to predict traffic by autonomous agents within a vehicle route planning system.

Estimate the traffic flow of a road segment by analyzing taxi trajectories. A recent study successfully uses the weather situations as supplementary information in traffic prediction model to enhance the prediction accuracy. In this work, any traffic prediction model can be used. In this work, the traffic parameter of interest is consisted of discrete classes, thus we treat traffic prediction as a classification problem. To be robust, different linear inference methods, including support vector machine (SVM), logistic regression (a.k.a. logit), and perception. The average accuracy of a 10-fold cross-validation is used to compute the accuracy
difference as shown in Figure. Our framework is compatible to various inference models to infer WTI.

**Fig 4:** Traffic prediction system

**KEY FACTOR VERIFICATION BY INDEX INFEERENCE**

The weather traffic index of one region can be inferred from the indices of its closely located (or adjacent) cells. The intuition is feasible because all the regions are connected by the road network, which can directly show the sensitivities of regions against weather. Based on the intuition, give a set of regional features \( F_r \), if the inference accuracy is satisfactory using \( F_r \) as input, it indicates that such set of regional features are the key factors the parent node \( g_u \) specifies the source cell, and the child node \( g_i \) is a set of observed cells which are closely located to \( g_u \). This model is not symmetric since \( g_i \rightarrow g_u \) may have a different probability comparing with \( g_u \rightarrow g_i \). The inference model can be any graphical classifier but use naive Bayes classifier, because the location closeness can be naturally considered by naive Bayes classifier. Since different cells have different numbers of neighbouring cells, it is hard to use other classifiers such as logistic regression, SVM, neural network, and random forest where the number of input features is fixed.

**MARGINAL DISTRIBUTION WEIGHT ESTIMATION OF REGIONAL FEATURES**

A marginal distribution describes the probability distribution of the regions contained in a similarity subset. Specifically in this paper, it describes the probability of one region being the index of \( i \) given one of its adjacent regions with index \( j \), if the two regions have a certain similarity. Given a set of regional features, some of them may have trivial impact to weather-traffic index, or are just noise. This requires us to test the weight of each regional feature through a feature selection process.

For example, Fisher score where features are scored by considering that features with high quality should assign similar values to instances in the same class and different values to instances from different classes; and Relief, which selects features to separate instance from different classes. In this case, a regional feature has nontrivial impact to weather traffic index. Let us remove this regional feature from the set of regional features.

The set of regional features is still the set of key factors which results in high overall accuracy. If not, it is a strong signal that the removed regional feature is very important; otherwise, it is less important. Use \( \delta(F_r) \) to denote the weight of the regional feature \( F_r \). The similarity of every two adjacent cells are recomputed in terms of the remained regional features, as well as the marginal distribution.

**Weight Estimation of Regional Features**

Given a set of regional features, some of them may have trivial impact to weather-traffic index, or are just noise. This requires us to test the weight of each regional feature through a feature selection process. There are many feature selection methods could be used in this paper. For example, Fisher score, where features are
scored by considering that features with high quality should assign similar values to instances in the same class and different values to instances from different classes; and Relief+, which selects features to separate instance from different classes suppose a regional feature has nontrivial impact to weather traffic index. Let us remove this regional feature from the set of regional features. the KFVII method is remaining set of regional features is still the set of key factors which results in high overall

accuracy. If not, it is a strong signal that the removed regional feature is very important; otherwise, it is less important. use δ(Fi r ) to denote the weight of the regional feature Fi r . Look closely, the similarity of every two adjacent cells are recomputed in terms of the remained regional features, as well as the marginal distribution. If the inference accuracy is increased more, the removed regional feature has more weight.

TRAVELING TIME PREDICTION ON SVM FOR WEATHER TRAFFIC INDEX

Travel-time calculation depends on vehicle speed, traffic flow and occupancy, which are highly sensitive to weather conditions and traffic incidents. These features make travel-time predictions very complex and difficult to reach optimal accuracy. Nonetheless, daily, weekly and seasonal patterns can still be observed at a large scale. For instance, daily patterns distinguish rush hour and late night traffic, weekly patterns distinguish weekday and weekend traffic, while seasonal patterns distinguish winter and summer traffic. The time-varying feature germane to traffic behaviour is the key to traffic-time modelling.

The main idea of the traffic forecasting is based on the fact that traffic behaviours possess both partially deterministic and partially chaotic properties. Forecasting results can be obtained by reconstructing the deterministic traffic motion and predicting the random behaviours caused by unanticipated factors. Suppose that currently it is time t. Given the historical data \( f(t-l), f(t-2), ..., f(t-n) \) at time \( t-l, t-2, ..., t-n \), we can predict the future value of \( f(t+1), f(t+2), ... \) by analyzing historical data set. Hence, future values can be forecasted based on the correlation between the time-variant historical data set and its outcomes.

Prediction Methodology and Error Measurements

Suppose the current time is \( t \), we want to predict \( y(t+l) \) for the future time \( t+l \) with the knowledge of the value \( y(t-n), y(t-n+1), ..., y(t) \) for past time \( t-n, t-n+1, ..., t \), respectively. The prediction function is expressed as:

\[
y(t+l) = \sum f(t, l, y(t+l), y(t)+l, ..., y(t))\]

The travel times of different prediction methods for departing from 7am to 10am during the last week between March 1 and March 4, 2018. Relative Mean Errors (RME) and Root Mean Squared Errors (RMSE) are applied as performance indices.

\[
RME = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{Y_i - Y_i'}{Y_i} \right|
\]

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left( \frac{Y_i - Y_i'}{Y_i} \right)^2}
\]

where \( Y_i \) is the observation value and \( Y_i' \) is the predicted value. Time-varying coefficients for travel time prediction. The estimated travel time \( T(t) \) and the snapshot travel time prediction \( T*(t, Δ) \) can be fitted to a linear curve as the relationship between \( T(t) \) and \( T*(t, Δ) \) since both are expected to slowly vary with respect to \( t \) and \( Δ \). Snapshot prediction is the future estimate of the travel time for a section of a freeway given that the speed with which the traffic traverses the freeway remains unchanged for the time. Since the relationship varies with time, the coefficients also vary with respect to time. Hence it is called a time-varying regression model.

Support Vector Regression Prediction Method

There are many parameters that must be set for travel-time prediction with support vector regression. We have tried several combinations, and finally chose a linear function as the kernel for performance comparison with \( \varepsilon=0.01 \) and \( C=1000 \).

Statistical Learning Theory

The statistical learning theory provides a framework for studying the problem of gaining knowledge, making predictions, making decisions from a set of data. In simple terms, it enables the choosing of the hyper plane space such a way that it closely represents the underlying function in the target space.

In statistical learning theory the problem of supervised learning is formulated as follows. We are given a set of training data \( \{(x_1,y_1),...,(x_i,y_i)\} \) in \( \mathbb{R}^n \times \mathbb{R} \) sampled according to unknown probability distribution \( P(x,y) \), and a loss function \( V(y,f(x)) \) that measures the error, for a given \( x \), \( f(x) \) is “predicted” instead of the actual value \( y \). The problem consists in finding a function \( f \) that minimizes the expectation of the error on new data that is, finding a function \( f \) that minimizes the expected error:

\[
\int V(y,f(x))P(x,y)dxdy.
\]

In statistical modelling we would choose a model from the hypothesis space, which is closest (with respect to some error measure) to the underlying function in the target space. More on statistical learning theory can be found on introduction to statistical learning theory.

Learning and Generalization

Early machine learning algorithms aimed to learn representations of simple functions. Hence, the goal of learning was to output a hypothesis that performed the correct classification of the training data and early learning algorithms were designed to find such an accurate fit to the data. The ability of a hypothesis to correctly classify data not in the training set is known as its generalization. SVM
performs better in term of not over generalization when the neural networks might end up over generalizing easily. Another thing to observe is to find where to make the best trade-off in trading complexity with the number of epochs. Firstly working with neural networks for supervised and unsupervised learning showed good results while used for such learning applications. MLP’s uses feed forward and recurrent networks. Multilayer perception (MLP) properties include universal approximation of continuous nonlinear functions and include learning with input-output patterns and also involve advanced network architectures with multiple inputs and outputs. These are simple visualizations just to have a overview as how neural network looks like. There can be some issues noticed. Some of them are having many local minima and also finding how many neurons might be needed for a task is another issue which determines whether optimality of that NN is reached. Another thing to note is that even if the neural network solutions used tends to converge, this may not result in a unique solution. Now let us look at another example where we plot the data and try to classify it and we see that there are many hyper planes which can classify it. But which one is better?

![Fig 5: Data classification](image)

There are many hyper planes which can be fit in to classify the data but which one is the best is the right or correct solution. From above illustration, there are many linear classifiers (hyper planes) that separate the data. However only one of these achieves maximum separation. The reason we need it is because if we use a hyper plane to classify, it might end up closer to one set of datasets compared to others and we do not want this to happen and thus we see that the concept of maximum margin classifier or hyper plane as an apparent solution. The next illustration gives the maximum margin classifier example which provides a solution to the above mentioned problem.
Expression for Maximum margin is given as (for more information visit [4])

\[
\text{margin} = \arg \min_{\mathbf{x} \in \mathcal{D}} d(\mathbf{x}) = \arg \min_{\mathbf{x} \in \mathcal{D}} \frac{|\mathbf{x} \cdot \mathbf{w} + b|}{\sqrt{\sum_{i=1}^{d} w_i^2}}
\]

The above illustration is the maximum linear classifier with the maximum range. In this context it is an example of a simple linear SVM classifier. Another interesting question is why maximum margin? There are some good explanations which include better empirical performance. Another reason is that even if we’ve made a small error in the location of the boundary this gives us least chance of causing a misclassification. The other advantage would be avoiding local minima and better classification. Now we try to express the SVM mathematically and for this tutorial we try to present a linear SVM. The goals of SVM are separating the data with hyper plane and extend this to non-linear boundaries using kernel trick. For calculating the SVM we see that the goal is to correctly classify all the data. For mathematical calculations we have,

[a] If \( Y_i = +1 \): \( w \mathbf{x}_i + b \geq 1 \)
[b] If \( Y_i = -1 \): \( w \mathbf{x}_i + b \leq 1 \)
[c] For all \( i \): \( y_i (w \mathbf{x}_i + b) \geq 1 \)

**SVM for Classification**

SVM is a useful technique for data classification. Even though it’s considered that Neural Networks are easier to use than this, however, sometimes unsatisfactory results are obtained. A classification task usually involves with training and testing data which consist of some data instances. Each instance in the training set contains one target values and several attributes. The goal of SVM is to produce a model which predicts target value of data instances in the testing set which are given only the attributes.

Classification in SVM is an example of Supervised Learning. Known labels help indicate whether the system is performing in a right way or not. This information points to a desired response, validating the accuracy of the system, or be used to help the system learn to act correctly. A step in SVM classification involves identification as which are intimately connected to the known classes. This is called feature selection or feature extraction. Feature selection and SVM classification together have a use even when prediction of unknown samples is not necessary. They can be used to identify key sets which are involved in whatever processes distinguish the classes.

**SVM for Regression**

SVMs can also be applied to regression problems by the introduction of an alternative loss function. The loss function must be modified to include a distance measure. The regression can be linear and non linear. Linear models mainly consist of the following loss functions, e-intensive loss functions, quadratic and Huber loss function. Similarly to classification problems, a non-linear model is usually required to adequately model data. In the same manner as the non-linear SVC approach, a non-linear mapping can be used to map the data into a high dimensional feature space where linear regression is performed. The kernel approach is again employed to address the curse of dimensionality. In the regression method there are considerations based on prior knowledge of the problem and the distribution of the
noise. In the absence of such information Huber’s robust loss function, has been shown to be a good alternative.

Support Vector Machines
The Support Vector Machine (SVM) is a powerful machine learning tool based on firm statistical and mathematical foundations concerning generalization and optimization theory. It offers a robust technique for many aspects of data mining including classification, regression, and outlier detection. SVM is based on Vapnik’s statistical learning theory and falls at the intersection of kernel methods and maximum margin classifiers. Support vector machines have been successfully applied to many real-world problems such as face detection, intrusion detection, handwriting recognition, information extraction, and others.

Support Vector Machine is an attractive method due to its high generalization capability and its ability to handle high-dimensional input data. Compared to neural networks or decision trees, SVM does not suffer from the local minima problem, it has fewer learning parameters to select, and it produces stable and reproducible results. If two SVMs are trained on the same data with the same learning parameters, they produce the same results independent of the optimization algorithm they use. However, SVMs suffer from slow training especially with non-linear kernels and with large input data size. Support vector machines are primarily binary classifiers. Extensions to multi-class problems are most often done by combining several binary machines in order to produce the final multi-classification results. The more difficult problem of training one SVM to classify all classes uses much more complex optimization algorithms and are much slower to train than binary classifiers. In the following sections, we present the SVM mathematical foundation for the binary classification case, then discuss the different approaches applied for multi-classification.

Proposed SVM based Algorithm
Step 1: In first Step dataset a training set is implemented to build up a model, while a test (or validation) set is to validate the model built. Data points in the training set are excluded from the test (validation) set. Usually a dataset is divided into a training set, a validation set in each iteration, or divided into a training set, a validation set and a test set in each iteration. Training set: a set of examples used for learning: to fit the parameters of the classifier. In the SVM case, we would use the training set to find the “optimal” Support Vectors. Validation set: a set of examples used to tune the parameters of a classifier. For SVM case, we would use the validation set to find the “optimal” number of support vectors or determine a stopping point for the algorithm. Test set: a set of examples used only to assess the performance of a fully-trained classifier. In the SVM case, we would use the test to estimate the error rate, FP rate or TP rate after we have chosen the final model.

Step 2: Scaling Feature Vectors: In this process the feature vectors are scaled or transformed into numeric vectors as the dataset contains various strings like 0 x<100 conversion or scaling of these features is required for SVM as the SVM can only work with numeric inputs.

Step 3: RBF Kernel (or Kernel Selection): The Gaussian RBF kernel is very popular and makes a good default kernel especially in absence of expert knowledge about data and domain because it kind of subsumes polynomial and linear kernel as well. Linear Kernels and Polynomial Kernels are a special case of Gaussian RBF kernel. Gaussian RBF kernels are non-parametric model which essentially means that the complexity of the model is potentially infinite because the number of analytic functions are infinite. If you see it from the point of view of polynomial functions, it essentially is infinite polynomial kernel. As Gaussian kernels including RBF are universal kernels i.e. their use with appropriate regularization guarantees a globally optimal predictor which minimizes both the estimation and approximation errors of a classifier. Here, approximation error refers to the error incurred by limiting the space of classification models over which search is performed, and estimation error refers to error in estimation of the model parameters. Also the credit dataset has the property of Exponential Gaussian curves which make the decision of choosing RBF function.

Step 4: Cross Validation: Cross-validation sometimes is a model validation technique for assessing how the results of a statistical analysis will generalize to an independent data set. It is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice. In a prediction problem, a model is usually given a dataset of known data on which training is run (training dataset), and a dataset of unknown data (or first seen data) against which the model is tested (testing dataset). The goal of cross validation is to define a dataset to “test” the model in the training phase (i.e., the validation dataset), in order to limit problems like over fitting, give an insight on how the model will generalize to an independent dataset (i.e., an unknown dataset, for instance from a real problem), etc. KFOLD analysis with 10 folds in RBF SVM was used for German credit dataset.

Step 5: SVM Training Model: The input feature data is fed into SVM Model for training and testing, after training the SVM Model is achieved which is used for prediction from the data. Matrix of training data, where each row corresponds to an observation or replicate, and each column corresponds to a feature or variable. For performance assessment, we use a test dataset with much lower fraud rate (0.5%) than in the training datasets with different levels of under sampling. This helps provide an indication of performance that may be expected when models are applied for fraud detection where the proportion of fraudulent transactions are typically low.

IMPLEMENTATION ENVIRONMENT
A weather-traffic index (WTI) system which mainly aim to fulfill two tasks. to set up a weather-traffic index throughout a city, which indicates the impact of weather to traffic from light to heavy, the second is to reveal the key factors behind the weather-traffic index throughout the city.
and their relative weights. Although there are many existing traffic prediction and measurement works as introduc- 

es in related works, they mainly focus on the analysis of road 

segments; on the contrary, the weather-traffic correlation 

throughout a city and conduct analysis of the key factors 

behind the regions whose traffics are highly influenced by 

inclement weather.

Weather forecasting entails predicting how the 

present state of the atmosphere will change. Present weather 

conditions are obtained by ground observations, 

observations from ships and aircraft, radio sound, Doppler 

radar, and satellites. This information is sent to 

meteorological centres where the data are collected, 

analysed, and made into a variety of charts, maps, and 

graphs. Modern high-speed computers transfer the many 

thousands of observations onto surface and upper-air maps. 

Computers draw the lines on the maps with help from 

meteorologists, who correct for any errors. A final map is 
called an analysis. Computers not only draw the maps but 
predict how the maps will look sometime in the future. The 
forecasting of weather by computer is known as numerical 
weather prediction. Weather forecasting has been one of the 
most scientifically and technologically challenging problems 
around the world in the last century. This is due mainly to 
two factors: first, it’s used for many human activities and 
secondly, due to the opportunism created by the various 
technological advances that are directly related to this 
congress research field, like the evolution of computation 
and the improvement in measurement systems. To make an 
accurate prediction is one of the major challenges facing 
meteorologist all over the world. Since ancient times, 
weather prediction has been one of the most interesting and 
fascinating domain. Scientists have tried to forecast 
meteorological characteristics using a number of methods, 
some of these methods being more accurate than others To 
predict the weather by numerical means, meteorologists 
have developed atmospheric models that approximate the 
atmosphere by using mathematical equations to describe 
how atmospheric temperature, pressure, and moisture will 
change over time. The equations are programmed into a 
computer and data on the present atmospheric conditions are 
folded into the computer. The computer solves the equations to 
determine how the different atmospheric variables will 
change over the next few minutes.

DATA PREPARATION

The road networks in the city of interest is 
partitioned into cells via Voronoi diagram where the seeds 
are road intersections. For each cell, the traffic parameters 
are extracted from taxi trajectories and the regional features 
are collected. The weather information of the same period of 
time is also collected.

WEATHER-TRAFFIC INDEX ESTABLISHMENT

The weather-traffic index is established for each 
cell by analyzing traffic data and weather data. In specific, 
given a cell g, the weather traffic index $\rho(g)$ is a value 
indicating the extent to which the traffic parameter in g is 
affected by weather.

FACTOR ANALYSIS

The input includes the established weather-traffic 
index and the regional features. The aim is to identify which 
regional features make traffic in cells vulnerable to 
inclement weather. In particular, the weights of regional features 
are quantitatively measured.

METHODOLOGY

1. The first is to set up a weather-traffic index 
throughout a city, which indicates the impact of 
weather to traffic from light to heavy.
2. The second is to reveal the key factors behind the 
weather-traffic index throughout the city and their 
relative weights.
3. It is mainly focus on the analysis of road segments; 
on local traffic-weather sensitivity throughout a 
city and the investigation to reveal the key.
4. Voronoi cell is the unit region of weather-traffic 
index. For each cell, traffic and weather of a long 
period of time are analyzed to decide the weather- 
traffic index.
5. Traffic parameter can be classified in terms of one 
of the following: quantity measures, e.g., “how 
much or at what rate is traffic moving or waiting to 
move?”, quality assessment measures.
6. Time mean speed (also called average speed) is 
used as the traffic parameter, which is defined as 
the arithmetic mean of individual spot speeds that 
are recorded over a selected time period.
7. Support vector machine SVM is both suitable for 
time series prediction and weather-traffic 
inference. Where logistic regression and perception, 
both of them are popular linear models, to verify 
the output of support vector machine. A special 
value is assigned to the cell.

The index throughout a city and analysis of key factors 
behind the index block diagram crash likelihood prediction 
model based on both real-time traffic flow variables 
measured through series of underground sensors and the 
rain data collected at weather stations in order to alarm 
potential crash occurrence in advance.

SYSTEM SPECIFICATION 
APPLICATION SPECIFICATION

Java language is a new computer 
programming language developed by Sun Microsystems. 
Java has a good chance to be the first really successful new 
computer language in several decades. Advanced 
programmers like it because it has a clean, well-designed 
definition. Business likes it because it dominates an 
important new application, Web programming.

Java has several important features:
A Java program runs exactly the same way on all computers. Most other languages allow small differences in interpretation of the standards.

It is not just the source that is portable. A Java program is a stream of bytes that can be run on any machine. An interpreter program is built into Web browsers, though it can run separately. Java programs can be distributed through the Web to any client computer.

- Java applets are safe. The interpreter program does not allow Java code loaded from the network to access local disk files, other machines on the local network, or local databases.

A group at Sun reluctantly invented Java when they decided that existing computer languages could not solve the problem of distributing applications over the network. C++ inherited many unsafe practices from the old C language. Basic was too static and constrained to support the development of large applications and libraries.

Today, every major vendor supports Java. Netscape incorporates Java support in every version of its Browser and Server products. Oracle will support Java on the Client, the Web Server, and the Database Server. IBM looks to Java to solve the problems caused by its heterogeneous product line. The Java programming language and environment is designed to solve a number of problems in modern programming practice. It has many interesting features that make it an ideal language for software development.

FEATURES OF JAVA

Sun describes Java as

- Simple
- Object-oriented
- Distributed
- Robust
- Secure
- Architecture Neutral
- Portable
- Interpreted
- High performance
- Multithreaded
- Dynamic.

Java is simple

What it means by simple is being small and familiar. Sun designed Java as closely to C++ as possible in order to make the system more comprehensible, but removed many rarely used, poorly understood, confusing features of C++. These primarily include operator overloading, multiple inheritance, and extensive automatic coercions. The most important simplification is that Java does not use pointers and implements automatic garbage collection so that we don't need to worry about dangling pointers, invalid pointer references, and memory leaks and memory management.

Java is object-oriented

This means that the programmer can focus on the data in his application and the interface to it. In Java, everything must be done via method invocation for a Java object. We must view our whole application as an object; an object of a particular class.

Java is distributed.

Java is designed to support applications on networks. Java supports various levels of network connectivity through classes in java. net. For instance, the URL class provides a very simple interface to networking. If we want more control over the downloading data than is through simpler URL methods, we would use a Reconnection object which is returned by a URL URL open Connection () method. Also, you can do your own networking with the Socket and Server Socket classes.

Java is robust

Java is designed for writing highly reliable or robust software. Java puts a lot of emphasis on early checking for possible problems, later dynamic (runtime) checking, and eliminating situations that are error prone. The removal of pointers eliminates the possibility of overwriting memory and corrupting data.

Java is secure

Java is intended to be used in networked environments. Toward that end, Java implements several security mechanisms to protect us against malicious code that might try to invade your file system. Java provides a firewall between a networked application and our computer.

FEATURES OF MY SQL

ROW VERSIONING BASED

ISOLATION LEVELS

This new database engine feature improves database read concurrency by reducing the amount of locks
being in user database. There are two versions of this feature (Both of which must be enabled at the database level) Read Committed isolation using Row versioning used at the individual statement level, and guarantees that the data is consistent for the duration of the statement.

ALWAYS ON AVAILABILITY GROUPS

This feature takes database mirroring to a whole new level with always on users will be able to fail over multiple databases in groups instead of individually. Also secondary copies will be readable and can be used for database backups. The big win is that users DR environment no longer needs to sit idle.

DEFINED SERVER TOOLS

DBAs always have the ability to create custom database role, but never server wide. For example, if the DBA wanted to give a development learn read write access to every database On a shared server, traditionally the only ways to do it were either manually or using undocumented procedures, neither of which were good solutions. Now the DBA can create a role, which has read write access on every DB on the server or any other custom server wide role.

ENHANCED AUDITING FEATURES

Audit is now available in all editions of SQL Server, additionally; users can define custom audit specifications to write custom events into the audit log new filtering features give greater flexibility in choosing which events to write to the log.

SEQUENCE OBJECTS

For those folks who have worked with oracle. This has been a long requested feature. A sequence is just an object that is a counter a good example of its use would be to increment values in a table, based a trigger. SQL has always had similar functionality with identify columns but now this a discrete object query optimization and it’s performed by the query processor itself.

SYSTEM TESTING

After the source code has been completed, documented as related data structures. Completion of the project has to undergo testing and validation where there is subtitle and definite attempt to get errors. The project developer treats lightly, designing and execution of the project test that will demonstrates that the program works rather than uncovering errors, unfortunately errors will be present and if the project developer doesn’t found errors, the user will find out.

The project developer is always responsible for testing the individual units i.e. modules of the program. In many cases, developer should conduct the integration testing i.e. the testing step that leads to the construction of the complete program structure.

The software, which has been developed, has to be tested to prove its validity. Testing is considered to be the least creative phase of the whole cycle of system design. In the real sense it is the phase, which helps to bring out the creativity of the other phases, and makes it shine.

The “Weather Traffic index” was tested along the following guidelines to prove its validity. It was tested using the following two techniques of software testing.

White Box Testing

- By using this technique it was tested that all the individual logical paths were executed at least once.
- All the logical decisions were tested on both their true and false sides.
- All the loops were tested with data in between the ranges and especially at the boundary values.

Black Box Testing

- By the use of this technique the missing functions were identified and placed in their positions.
- The errors in the interfaces were identified and corrected.
- This technique was also used to identify the initialization and termination errors and correct them.

Testing Methodologies

Testing is the process where the test data is prepared and is used for testing the modules individually and later the validation given for the fields. Then the system testing takes place which makes sure that all components of the system property functions as a unit. Testing is best performed when user departments are asked to assists in identifying all possible situations that might arise. The test data should be chosen such that it passed through all possible condition. Actually testing is the state of implementation which aimed at ensuring that the system works accurately and efficiently before the actual operation commence.

The following is the description of the testing strategies, which were carried out during the testing period.

Code Testing
On the execution of the program for a set of data are repeated. Thus the code was exhaustively checked for all possible correct data and the outcomes were also checked.

**Module Testing**

To locate errors, each module is tested individually. This enables us to detect errors and correct it without affecting any other modules. Whenever the program is not satisfying the required function, it must be corrected to get the required result. Thus all the modules are individually tested from bottom up starting with the smallest and lowest modules and proceeding to the next level.

Each module in the system is tested separately. For example company module is tested separately. This module is tested with different month data and the result of the test is compared with the results that are prepared manually. The comparison shows that the results prepared by the new system is absolutely correct than the existing system.

**Integration Testing**

After the module testing, the integration testing is applied. When linking the modules there may be chance to occur errors. These errors are corrected by using this testing. In this system all modules are connected and tested. The testing result are very correct.

**Acceptance Testing**

When that user finds no major problems with its accuracy, the system passes through a final acceptance test. This test confirms that the system needs the original goals, objectives and requirements established during analysis without actual execution which elimination wastage of time and money acceptance tests on the shoulders of users and management, it is finally acceptable and ready for the operation.

**EXPECTED OUTCOME**

**EVALUATION METRICS**

The performance and accuracy are calculated based on the following measures:

- **Accuracy** = (TN+TP) / (TN+TP+FN+FP)
- **Precision** = (TP) / (TP+FP)
- **Recall** = TP / (TP+FN)

- **TN**: True Negative, Legitimate predicted as Legitimate
- **TP**: True Positive, Spam predicted as Spam
- **FP**: Legitimate predicted as Spam
- **FN**: Spam predicted as Legitimate

**Expected reciprocal rank (ERR)** is defined as the expected reciprocal rank evaluates each inference result which indicates the likelihood for this cell to take each index value.

**SUMMARY**

This chapter describes about the evaluation metrics. That metrics used to estimate the performance and accuracy of the expected outcome that the weather-trafic index done for the user achieves high accuracy. In information retrieval, precision (the retrieved instances that are relevant), while recall (the fraction of relevant instances that are retrieved).

**CONCLUSION**

The impact of weather to traffic from few locations to all road networks throughout a city, more importantly, the regional features leading to the vulnerability of traffic in local areas to inclement weather are systematically revealed for the first time. The regional weather-trafic indices extracted have been validated to be surprisingly consistent with real world observations. Detect weather patterns which are consistent with global climate change on weather patterns. Detect rare weather patterns in greater detail by adjusting the resolution of the proposed detection model.

This method presents an efficient data processing technique used to enable the available data proper for the knowledge discovery as if the available metrological data is not proper the detection can be improper which leads to improper prediction of climate change. The goal of the proposed method is to implement an effective, efficient and adaptive rare pattern detection technique consisting of four parts missing data recognition, data pre processing, initial clustering and adaptive clustering for dynamic detection.

**REFERENCES**

1. Y. Ding, Y. Li, K. Deng, H. Tan, M. Yuan, and L. M. Ni, “Dissecting regional weather-trafic sensitivity throughout a city,” in 15th IEEE International Conference on Data Mining, ICDM 2015, Atlantic City, NJ, USA, November 14-17, 2015, pp. 739–744.
2. Y. Zheng, L. Capra, O. Wolfson, and H. Yang, “Urban computing: Concepts, methodologies, and applications,” ACM Transaction on Intelligent Systems and Technology, 2014.
3. A. Saegusa and Y. Fujiwara, “A study on forecasting road sensitivity based on weather and road surface data,” IEICE Transactions, vol. 90-D, no. 2, pp. 509–516, 2007.
4. M. A. Abdel-Aty and R. Penmanaboina, “Calibrating a real-time traffic crash-prediction model using archived weather and its traffic data,” IEEE Transactions on...
Intelligent Transportation Systems, vol. 7, no. 2, pp. 167–174, 2006.
[5] S. Dunne and B. Ghosh, “Weather adaptive traffic prediction using neurowavelet models,” IEEE Transactions on Intelligent Transportation Systems, vol. 14, no. 1, pp. 370–379, 2013.
[6] M. J. Koetse and P. Rietveld, “The impact of climate change and weather on transport: An overview of empirical findings,” Transportation Research Part D: Transport and Environment, vol. 14, no. 3, pp. 205–221, 2009.
[7] . Talking about smart transportation from the rainstorm of Beijing, “http://info.secu.hc360.com/2012/07/300822649026.shtml.”
[8] Y. Zheng, F. Liu, and H.-P. Hsieh, “U-air: when urban air quality inference meets big data,” in KDD, 2013, pp. 1436–1444.
[9] S. Liu, Y. Liu, L. M. Ni, J. Fan, and M. Li, “Towards mobility-based clustering,” in KDD, 2010, pp. 919–928.
[10] J. Yuan, Y. Zheng, and X. Xie, “Discovering regions of different functions in a city using human mobility and pois,” in KDD, 2012, pp. 186–194.
[11] F. Zhang, D. Wilkie, Y. Zheng, and X. Xie, “Sensing the pulse of urban refueling behavior,” in UbiComp, 2013, pp. 13–22.
[12] K. Zheng, Y. Zheng, N. J. Yuan, and S. Shang, “On discovery of gathering patterns from trajectories,” in ICDE, 2013, pp. 242–253.
[13] L. A. Tang, Y. Zheng, J. Yuan, J. Han, A. Leung, C.-C. Hung, and W.-C. Peng, “On discovery of traveling companions from streaming trajectories,” in ICDE, 2012, pp. 186–197.
[14] Y. Zheng, “Methodologies for cross-domain data fusion: An overview,” IEEE Trans. Big Data, vol. 1, no. 1, pp. 16–34, 2015.
[15] F. Aurenhammer, “Voronoi diagrams - a survey of a fundamental geometric data structure,” ACM Comput. Surv., vol. 23, no. 3, pp. 345–405, 1991.