Transportation Domain

New York City Traffic

Introduction: Big data is a popular term used to describe the exponential growth and availability of both structured and unstructured data. And big data may be as important to business – and society – as the Internet has become. More data may lead to more accurate analyses that leads to more confident decision making. And better decisions can mean greater operational efficiencies, cost reductions and reduced risk [1].

Big data is defined by 3 distinct features.

1. **Volume**: This describes the increase in data volume. Data is getting collected at a very fast pace and as technology has improved the cost to store this data has highly reduces. Hence it's economical to store large amounts of data.
2. **Variety**: The data getting collected or generated each day is present in various forms. It can be either text files, images, audio clips, video clips, emails etc., managing, merging and analyzing such vast types of data is highly challenging.
3. **Velocity**: Data is entering the system and also being generated at an unprecedented rate. There is a high need to quickly react to such data. This is challenging for most of the organizations.

There are several features of big data as stated by different organizations, however the most distinct features are the ones stated above.

Transportation Domain: New York is one of the busiest cities in the world. It is said to be the city that never sleeps. Entry into New York City (Manhattan) is either through bridges or tunnels. Most of the time the roads and entry points into NYC are highly congested. These roadways are managed by several authorities like Manhattan Port Authority, City Corporation, Metropolitan Transportation Authority. All these governing bodies have their own sets of rules and regulations to control the congestion. One of the famous methods used by them are collecting road tolls from travelers.

On an average there are around 254808 vehicles per day. The average speed of each vehicle is around 13.3 mph [2]. So one can very well imagine the snail pace at which the traffic moves. So the commuters end up spending almost 3 – 4 hours on a per day basis on the road. It is also observed that almost 10 – 20 % of the vehicles are looking for parking spots which make them travel around the same block in a slow speed in order to find an empty spot [3]. So it is apt to say that NYC needs a good solution to traffic and parking issues.

Problems: Peak traffic on the road leads to traffic congestion. This in turn leads to a couple of problems that need to be addressed soon. Traffic congestion causes delays. During the morning rush hour congestion not only causes delay but also causes a lot of stress as people are worried

that they will be late for work [4]. Again at the end of the day evening rush hour is a major headache as the work day is done and people wish to go home and relax and the traffic prevents it [4]. The travel time cannot be estimated beforehand due to peak traffic. Instead travelers end up being late for work or engagements which causes a lot of discomfort [4].Road rage is also common when there is a lot of congestion. Road rage results from stress and frustration. It causes a ripple effect among travelers and also causes a situation that is of high risk to himself and to the people around him [4].Traffic congestion also affects emergency vehicles as is it unable to respond on time [4].

Solutions Undertaken: Many roadways in NYC have proposed to collect special tolls called 'Congestion Pricing' [5], which is a special fare during peak hours. This prompts travelers to avoid travelling during peak hour instead travel during less congested traffic hours. Another solution would be to build more roads. However this is not a feasible solution. It is not possible to build roads in a short time span and also more reads will only lead to more traffic.

There has been a lot of research going on to find good solutions to problems caused by traffic congestion. Organizations are making use of traffic data to analyze traffic patterns and come up with solutions. A few leading research and the resulting solutions have been listed below.

Big Data Research:

1. **United Parcels Services (UPS):** UPS is one of the biggest shipping companies in the world. It was founded in 1907 and on a daily basis they make more than 16 million shipments to over 8.8 million customers globally. In addition, they receive on average 39.5 million tracking requests from customers per day. They employ in the US alone over 55,000 drivers and they have more than 100,000 vehicles globally. They gather data at every possible moment and already store over 16 petabytes of data. They have been working with Big Data for a very long time already [6].
   UPS carries out a lot of analyses on data collected and hence is able to provide unique services to the customers which in turn benefits the environment and brings profit to the company. One such project is Orion (On-Road Integrated Optimization and Navigation).Its objective is to provide all drivers with advanced mathematical models that provide additional optimization and navigational capabilities to make the drivers more efficient. And more efficient drivers mean saving a lot of money [6]. But UPS does not only apply Big Data analytics to optimize their routings. They have also developed algorithms that predict the maintenance requirements of their delivery trucks. With so many packages delivered daily, an unplanned breakdown of a truck can have serious consequences resulting in late packages and angry customers [6].
   Predictive maintenance helps UPS save millions in preventing unnecessary and unplanned maintenance. The massive amounts of data created by UPS are prescriptive analytics to handle the ever-changing operational landscape they have to deal with. All the different data feeds, optimized and changed routings and customer delivery preferences require powerful analytics to optimize their daily activities. Their efforts have paid off, because

thanks to all the data collected and the massive investments in Big Data, UPS has become the largest, and they claim also the most efficient, package handling company in the world [6].

ORION: Orion is the brain child of UPS. It's a route optimization software that took UPS almost 10 years to develop and test. ORION uses expansive fleet telematics and advanced algorithms to gather and calculate countless amounts of data to provide UPS drivers with optimized routes.

One of the algorithms used by ORION is the nearest Neighbor algorithm which is used to solve the travelling salesman problem. The pseudo code for the algorithm is given below:

Problem: Given a list of cities and the distances between each pair of cities, what is the shortest possible route that visits each city exactly once and returns to the origin city?

Approximate Solution: Nearest Neighbor Algorithm [8]

1. *V = {1, ..., n-1} // Vertices except for 0.*
2. *U = {0} // Vertex 0.*
3. *while V not empty*
4. *u = most recently added vertex to U*
5. *Find vertex v in V closest to u*
6. *Add v to U and remove v from V.*
7. *endwhile*
8. *Output vertices in the order they were added to U*

Each business day, UPS drivers deliver between 125 and 175 packages. The number of route combinations a driver can make is far greater than the number of nanoseconds the earth has existed. To ensure UPS drivers use the most optimized delivery routes in regard to distance, fuel and time, UPS developed On-Road Integrated Optimization and Navigation (ORION) [7].

The ORION program is the result of a long-term operational technology investment and commitment by UPS that will be more than a decade in the making from the initial development of the algorithm to full deployment to nearly all 55,000 routes in the North American market by 2017. 2013 marked the first major ORION deployment by a team of 500 dedicated resources to rollout ORION to 10,000 UPS routes. As results are exceeding expectations, in 2014 UPS increased the ORION team to 700 to speed deployment to 45% of US routes by the end of the year [7].

The origin of ORION In 2008, UPS piloted telematics technologies on delivery trucks to gather data to understand where efficiencies can be improved. By installing GPS tracking equipment and vehicle sensors, combined with a driver's handheld mobile device, UPS could capture data related to vehicle routes, as well as the amount of time vehicles idled, and even whether drivers were wearing seatbelts[7].

While implementing solutions to gather the necessary data, UPS was developing a complex algorithm that could quickly solve complex routing problems. The resulting algorithm includes about 1000 pages of code and turns the captured data into instruction to optimize driver's routes [7].

Today, ORION can solve an individual route in seconds and is constantly running in the background evaluating routes before drivers even leave the facility. This level of route evaluation conducted through the ORION program requires extensive hardware and architectural provisions. Running on a bank of servers in Mahwah, NJ, ORION is constantly evaluating the best way for a route to run based on real-time information. While most of America is sleeping, ORION is solving tens of thousands route optimizations per minute.

In addition to architectural enhancements, the drivers' Delivery Information Acquisition Device (DIAD) is enhanced to serve as the tool for communicating optimized routes to drivers while on the road [7].

Results with ORION

With 10,000 routes optimized with ORION, UPS saves more than 1.5 million gallons of fuel and reduces carbon dioxide emissions by 14,000 metric tons. Initial results show miles reduced with each route using ORION and a reduction of just one mile per driver per day over one year can save UPS up to $50 million [7].

ORION also benefits customers as it will enable more personalized services, even on peak business days. This includes the UPS My Choice® service, which allows consumers to have online and mobile access to see their incoming UPS home deliveries, and enables them to actively choose delivery preferences, reroute shipments and adjust delivery locations and dates as needed. Currently millions of customers take advantage of the UPS My Choice service, and ORION technologies will continue to make possible even more personalized services, with international on the future roadmap [7].

Data Used:

- UPS uses a wide variety of data to reduce the environmental impact of their business
- They combine data from engines that provide UPS with insights in performance and condition of their vehicles
- Speed, number of stops, mileage and miles per gallon etc. are measured.
- They use GPS data that captures driver behavior and safety habits.
- GPS data to find out vehicle density on the roads.
- Sensors in vehicles report data on emissions and fuel consumptions

Long story short: they collect data at every possible moment.

Outcome:

Reduction of 6 to 8 miles driven per route resulting in lower fuel use and lower fuel emissions

- UPS saved 3 million gallons of fuel during its testing of the program from 2010-2012
- The 1.5 million gallons saved this year will also cut 14,000 metric tons of carbon dioxide emissions.
- NO left turns -left turns increased the company's carbon footprint, because of the gas trucks wasted and the emissions they created while waiting for traffic lights to change, by mapping routes that avoid left turns, the company saved 98 million minutes of idling time in 2011 alone, and the mile-reducing feature created a savings of nine million gallons of fuel.

Challenges:

The main challenge faced by UPS was the time taken to build ORION, also it was pretty difficult to get the UPS drivers to accept ORION. UPS drivers who travelled in the delivery route for many years, where not very happy to accept ORION.

2. **The worst place to park in NYC is discovered using Big Data**[9]

Data scientist Ben Wellington used public data collected by the city to make a positive change, and also stressed that further data sharing can benefit all metropolitan cities.
Data Used: GPS – acquired data of the average taxi speed in NYC throughout the day.
Analysis: He wanted to analyze when exactly rush hour started and ends in NYC.

Outcome:

- Between 8:30 AM to 6:30 PM , traffic moved at a pace of 11.5 mph
- Entire day the traffic moved at the same pace.

Additional Outcome: He found out the worst place to park in NYC. He analyzed the parking tickets that were issued in various locations. He found out that the most ticketed spot was near a fire hydrant where almost $33,000 was collected annually. With such an astonishing discovery, Wellington decided to visit the hydrant and was surprised to find misleading marks near the hydrant. Driver easily mistook it for a legitimate parking space, and for five straight years the NYPD would ticket them [9]. Wellington then wrote to traffic officials and informed them of the error that was causing drivers to be unfairly ticketed. A few weeks later the city repainted the area to clearly mark it as an illegal parking spot [9].

3. **IBM- Traffic management for a smarter planet**

IBM has proposed the idea on creating Intelligent Transportation which provides traffic analysis and prediction capabilities and a comprehensive, scalable platform for traffic management solutions. It aggregates data from multiple devices that identify and measure traffic speed and volume on city roads. Giving you near real-time citywide visibility into traffic conditions, it can capture data from disparate sources, such as

cameras, radar and under-road loop detectors, as well as systems based on Bluetooth or mobile phone technology[11].

Intelligent transportation collects data from various devices that identify and measure traffic on city roads. It performs analysis on the data and provides near real time information about traffic conditions.

The predictive analytics capability in the intelligent transportation solutions leverages historical and real-time traffic data to predict future traffic conditions. It adopts a system-wide analysis method, so the solutions can offer industry-leading accuracy when predicting conditions such as traffic speed and volume, up to an hour in advance [11].

Outcome:

- Analyze data to find out about traffic accident patterns and behavior of traffic during such incidents.
- Predict traffic speed and pattern up to an hour based on real time and historical data.
- Increase situation awareness.

4. **Boston Uses Big Data to Solve Traffic Jams**[12]

Driving in Boston is a challenge. Off late Boston has started analyzing big data so that commuters find their travel less stressful.

Problem: Commonwealth Ave in Boston is always congested with parked cars that cause the three lane road to convert into a two lane road often during rush hour.

Analysis: Uses Data compiled from Waze and other sources to deploy bicycle cops to ticket and/or tow cars double parked on Commonwealth Ave., [12]

Waze is a google owned smartphone app that crowd -sources traffic updates from some 450,000 local users. The city also is working with Uber to ease traffic woes by leveraging ride data [12].

The combination of data sources and street cameras means that officials in the city's traffic nerve center can work behind the scenes to ease congestion [12].

All of the local Waze data feeds into the city's traffic management center where officials can manipulate traffic lights based on what's happening on the ground, and then compare resulting conditions with what has happened historically at that site over a period of time [12].

Conclusion and Proposal for final project: If major cities decide to analyze big data relating to traffic then they can easily come up with solution to ease the traffic congestion problems during rush hour. There is a lot of research going on in this field. Hence I would like to continue my final project using the NYC traffic data set to come up with useful analysis.

References

1. http://www.sas.com/en_us/insights/big-data/what-is-big-data.html
2. http://gothamist.com/2010/03/24/fun_facts_about_manhattan_traffic_c.php
3. http://tomvanderbilt.com/traffic/some-things-about-traffic-that-may-surprise-you
4. http://traveltips.usatoday.com/effects-traffic-congestion-61043.html
5. https://en.wikipedia.org/wiki/Congestion_pricing_in_New_York_City
6. https://datafloq.com/read/ups-spends-1-billion-big-data-annually/273
7. https://www.pressroom.ups.com/pressroom/ContentDetailsViewer.page?ConceptType=Factsheets&id=1426321616277-282
8. https://en.wikipedia.org/wiki/Nearest_neighbour_algorithm
9. http://abc7.com/traffic/big-data-helps-find-the-worst-place-to-park-in-nyc/537430/
10. http://www.ibm.com/smarterplanet/us/en/traffic_congestion/article/traffic-management-and-prediction.html
11. http://www.ibm.com/smarterplanet/kw/en/traffic_congestion/nextsteps/solution/N501656B00014S33.html
12. http://fortune.com/2015/05/21/boston-is-using-big-data-to-solve-traffic-jams/
13. http://www.wsj.com/articles/at-ups-the-algorithm-is-the-driver-1424136536
14. http://www.fastcompany.com/3004319/brown-down-ups-drivers-vs-ups-algorithm