Queens College

Cassandra Newcomer
BS in Computer Science
Saranjeet Bhamra
MA in Data Analytics
Karishma Muni
BS in Computer Science

CUNY-IBM Watson Case Competition
SPRING 2018

BUSINESS CASE

April 2018

GROUP NAME: MeTAnalyzer
TEAM NUMBER: 29
TEAM LEAD: Cassandra Newcomer
BUSINESS CASE TITLE: Social media derived real time predictive analytics for the NYC public transit system.
VIDEO: https://youtu.be/MlVjdJAlaYg
SITE: https://metanalyzer.github.io/
DOMAIN: City Services
SUMMARY: MTA related content, photos, and metadata from social media can be aggregated and analyzed via a combination of image recognition, sentiment analysis, and other NLP/machine learning technologies available via the IBM Watson API. Insights gained will be used to optimize the detection, prevention, and communication of emergent transit problems in real time.

1. EXECUTIVE SUMMARY

New York City's subway system includes 665 miles of track, 472 stations, and 27 subway lines. It also has the worst on-time performance of any major rapid transit system in the world (1). Citizens have increasingly begun to rely not on the MTA, but their fellow commuters to glean information about current service status. Via social media, commuters can communicate about delays and outages necessitating alternate routing options. However, this process consumes time and data in a place where internet access is intermittent and battery supply a precious, steadily depleting resource. We propose leveraging the IBM Watson platform to aggregate, analyze, and visualize trends in commuter experiences. This will enable city services and the city's commuter population to cooperatively improve the state of the New York City public transit experience.

2. PROBLEM DEFINITION

For over a century, public transit has been the predominant mode of transportation for low and middle income New Yorkers. Ridership levels have increased significantly, yet the quality of MTA service has entered a precipitous decline. Frustrated commuters are increasingly leaving New York for other cities, often citing the deteriorating public transit situation and unpredictable commutes as the reason for their departure. If this trend continues, the urban renewal New York has enjoyed for most of the last decade is likely to halt or perhaps even reverse trajectory. The lost income due to transit delays is already estimated to be close to $389,000,000 dollars a year (2), a number that is only expected to rise. Restoring faith in the New York City public transit system is perhaps the most critical challenge faced by the city in the current year.

3. RECOMMENDED SOLUTION

The MTA is currently defined in the public sphere by its inability to appropriately communicate service changes to straphangers. It is close to impossible to discern if a train is running or not before entering a station, as the digital billboards frequently display expected arrival times for trains that are not even running (3). This often leads to train platforms becoming dangerously overcrowded, as more and more people enter a station that no one is able to leave. As a result, commuters – who while stranded on non functional transit, are now almost universally equipped with smartphones – have taken to manually scrolling through relevant hashtags on social media to find and/or post information about whether or not a

train line is functioning, and if not, then why. Because the only reliable source of information is their fellow commuters, much trust has been lost in the system. However, these tweets about service status and photos of track fires are a critical source of realtime data. Aggregating this crowdsourced information could and should be automated. In the short term, it permits commuters the ability to easily halt or resume use of an impacted station. In the long term, the city government and the MTA can analyse the dataset for insight into broader trends. This will facilitate high resolution graphing of the public perception of the system versus its quantifiable runtime state at any given temporal point.

> Jemilah Magnusson of Harlem checks two separate apps to figure out which train to take. "It adds another layer of work that you have to do in a city that's very demanding. . .to also need to problem-solve for what's going to happen with the subway — it's tough." (3)

4. IMPLEMENTATION

Twitter is the medium of choice for realtime subway complaints. The location of a complaint may be derived from both the textual content, the hashtags, or even the geolocation, making it an excellent source of metadata. Using IBM Watson, we can build a full stack web application to monitor MTA relevant hashtags and track when, where, and for how long an unhappy commuter has been experiencing service interruptions. (Happy commuters may also be observed to determine which stations are functioning optimally.) After analysis, the results will be integrated into a realtime data feed taking two forms: a mobile application (for use by commuters), and a dashboard application for data visualization and analytics (for use by the city). For commuters, the mobile app will allow them to subscribe to individual stations and view the actual service status as experienced and reported by their fellow citizens. This will preserve commuter sanity and increase a sense of civic participation as commuters will no longer feel as if their tweets and photos are disappearing "into the void," but rather that they are helping to improve the situation and aid their fellow New Yorkers.

For the city, the IBM Watson platform used in tandem with the MeTAnalyzer visualization dashboard will enable the discovery of previously uncharted relationships between the public transit system and the public it serves. It may be used by itself or in combination with pre-existing data sources to track longterm trends and even perform predictive analysis of when and where future transit breakdowns are likely to occur. The

electro-mechanical aspects of the New York City public transit system have typically been the focus of most improvement intitiatives so far. However, these omit a key factor: the humans interacting with the system. Any mathematical modeling which fails to take this into account will be sharply limited in its predictive ability. The MeTAnalyzer platform is therefore unique in that it holds the potential to create a virtuous feedback cycle. Once it becomes possible to measure objectively if a given initiative has, in fact, increased commuter morale, the city can iterate far more effectively towards a permanent solution. If the MeTAnalyzer mobile app becomes popular amongst commuters, an opt-in metadata collection feature would also allow the MTA to finally have a method of tracking, modeling, and (eventually) influencing commuter behavior patterns to further streamline service.

5. ORGANIZATIONAL IMPACT

The subway is currently in such a fragile state that a single breakdown often leads to delays across multiple lines. Without effective backpressure routing, choke points rapidly become overwhelmed by the influx of commuters fleeing nonfunctional stations. This overcrowding subsequently causes additional breakdowns and delays. As the mobile app will permit commuters to reroute their itineraries in response to events, wear and tear on the system as a whole should be immediately reduced. Many departments, including first responders and TRANSCOM, should expect to see significant growth in efficiency. Service can be restored far more quickly when the NYPD and NYFD do not have hundreds of unhappy and confused commuters between them and the crisis awaiting them at the subway platform. With live photos and first person descriptions of the situation already collected for easy reference, terrorist attacks or track fires can be responded to at a scale appropriate to the problem. At the start of this year, nearly 10,000 subway delays per month had no known cause (5). The aggregated dataset will also allow for automated collection and categorization of events.

6. CONCLUSION

The IBM Watson platform holds the potential to promote seamless communication between the MTA, city services, and the New Yorkers who rely upon them for their daily commute. This will increase the quality of service and aid in reducing the logistical, financial, and emotional impact of the subway outages in the short term. It is also well positioned to help fix the problem more completely in the long term.

CITATIONS

1. "M.T.A. Delays: How Did the Subway Get So Bad?" The New York Times. Feburary 20, 2018. https://www.nytimes.com/2018/02/20/nyregion/mta-train-delays.html

2. "The Subway Is Being Fixed, but Can Riders Tell?" The New York Times. February 7, 2018. https://www.nytimes.com/2018/02/07/nyregion/subway-delays-cuomo-byford.html

3. "The MTA's New Arrival Clocks Don't Really Work." New York Post. January 3, 2018. https://nypost.com/2018/01/03/the-mtas-new-arrival-clocks-dont-really-work/

4. "Left in the Dark: How the MTA Is Failing to Keep Up With New York City's Changing Economy." New York City Comptroller. March 23, 2018. https://comptroller.nyc.gov/reports/left-in-the-dark-how-the-mta-is-failing-to-keep-up-with-new-york-citys-changing-economy/

5. "MTA Gave False Reasons for 10,000 Subway Delays That Originally Couldn't Be Explained." The New York Daily News. March 27, 2018. http://www.nydailynews.com/new-york/mta-gave-false-reasons-10-000-subway-delays-article-1.3897815