

CUNY-IBM Watson Case Competition 2018

Team Number: 29

Team Name: MeTAnalyzer

Business Case Title: Social media derived real time predictive analytics for the NYC public transit system.

Link to the 1-minute (Youtube / vimeo link): <https://youtu.be/JaRJ6RVnAp0>

Site: <http://metanalyzer.github.io>

Sector (Higher Education, or City Services): City Services

One-line about your project (Give us your best one-sentence pitch):

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.

Team Members : Cassandra Newcomer, Karishma Muni, Saranjeet Bhamra

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 instead their fellow commuters to glean information about transit 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. Leveraging a combination of IBM Watson's sentiment analysis, NLP, and machine learning capabilities to aggregate trends in commuter experiences will help 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 significant decline. Frustrated commuters are increasingly leaving New York for other cities, often directly 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 (3), 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

Presently, it is close to impossible to discern if a train is running or not before entering a station. The automatic billboards frequently display expected arrival times for trains that are not running, leading train platforms to become dangerously overcrowded as more and more people enter a station that no one is able to leave. The MTA is now notorious for its inability to communicate the current situation to straphangers. 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, what the reason might be for the delay. Because the only reliable source of information is their fellow commuters,

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.” (2)

much trust has been lost in the system. The tweets about service status and photos of things like 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 pause or resume patronization of an impacted station. In the long term, the city government and the MTA can analyse the data for insights and broader trends, as it will allow for high resolution graphing of the public perception of the system vs its quantifiable runtime state at any given temporal point.

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 (for use by the city). For commuters, the mobile app will allow them to subscribe to individual stations and view the realtime status as reported by their fellows so they know to reroute around impacted stations. 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 citizens. For the city, a data visualization dashboard can monitor the status in realtime and track longterm trends.

5. ORGANIZATIONAL IMPACT

Many departments, including first responders and TRANSCOM, will see an immediate 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 descriptions of the situation already aggregated for easy reference, terrorist attacks or track fires can be responded to at a scale appropriate to the problem. Using Watson and the MeTAnalyzer visualization dashboard, we can also discover relationships and previously overlooked common factors in predicting and ameliorating unexpected service changes, which can be used by itself or in combination with pre-existing data sources to perform predictive analysis of when and where future transit breakdowns are likely to occur. At the start of this year, nearly 10,000 subway delays per month had no known cause (4). The aggregated dataset will also allow for easy automated collection and categorization of events that can be used to identify probable causes.

The subway is currently in such a fragile state that a single breakdown often leads to delays on multiple unrelated lines as the failure rapidly propagates through the system. Due to the current lack of backpressure routing, choke points become overwhelmed by the influx of commuters fleeing nonfunctional lines, with this overcrowding being a frequently listed cause of further 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 greatly reduced. Currently, the MTA has some idea about the performance of the mechanical parts of the system in response to failure, but relatively little data about how the human parts react during the same event. MeTAnalyzer holds the potential of creating a virtuous feedback cycle, where the city can iterate effectively towards a permanent solution as it will be possible to measure objectively if a given initiative has, in fact, increased commuter morale. If the MeTAnalyzer 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 streamline service even further. Perhaps most critically, the city would realize substantial financial gains from the taxes paid by workers who do not lose their employment or leave the city due to our public transit failing to serve our public effectively.

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. February 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. “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/>
4. “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>