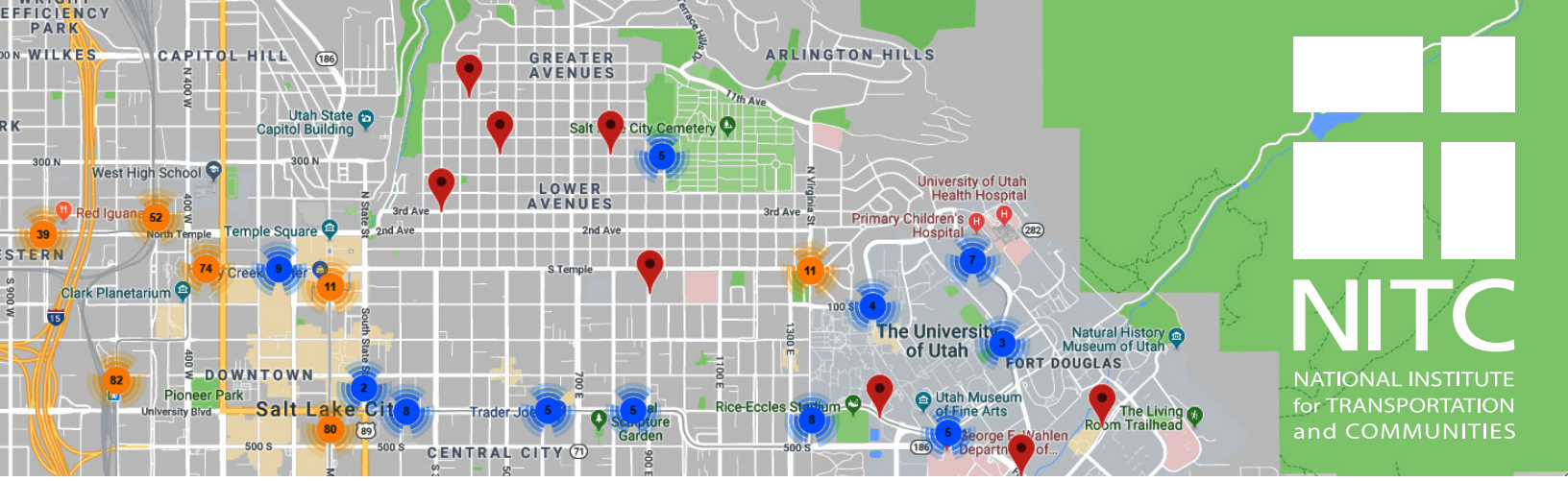

Leveraging Twitter and Machine Learning For Real-Time Transit Network Evaluation

Xiaoyue Liu, Ph.D., Ran Wei, Ph.D., Aaron Golub, Ph.D., Liming Wang, Ph.D.

With today's profusion of open data sources and real-time feeds, transit agencies have an unparalleled opportunity to leverage large amounts of data to improve transit service. Thanks to NITC researchers, there is now an open-source tool for that.

The new **Social-Transportation Analytic Toolbox (STAT)** for Transit Networks, developed by researchers at the University of Utah and Portland State University, is a dynamic platform that combines Twitter, general transit feed specification (GTFS), and census transportation planning products (CTPP)—in this case, job density data—to help agencies evaluate overall system performance and identify connectivity gaps. It can also act as a decision support tool for recommending service improvements.

THE SOCIAL-TRANSPORTATION ANALYTIC TOOLBOX (STAT)

The STAT is an open-source, publicly accessible toolbox with three components:

- Temporal distribution of transit stops' average travel times
- Transit stop positioning in Google Maps with geo-mapped tweets around that stop
- Overall transit access visualization at the TAZ (traffic analysis zone) level

The tool is a potential game changer. Agencies usually resort to annual surveys to obtain transit riders' opinions about service. This is a high-cost, low-resolution method that can reach only a limited sample of transit system users. The STAT tool can help transit providers see a more complete picture of the dynamic and complex interactions between riders and services.

WEIGHTED AVERAGE TRAVEL TIME FOR ALL TIMES OF DAY

Weighted average travel time (WATT), is a transit performance measure that weights travel times (from one stop to all possible stops) based on the attractiveness (potential opportunities) of destinations. In this particular case, travel time is weighted based on job density retrieved from the census data. The major drawback of past studies using WATT is that they all have the same missing piece: the variance in travel times at different times of day. Calculating WATT for all times of day, as this tool does, provides a comprehensive transit accessibility measure that captures the temporal variation in services.

LEVERAGING DATA IN TRANSPORTATION

Forward-thinking transportation analytics has started to realize the advantages of using the explosion of data to manage mobility. For example, the city of Los Angeles partnered with Google Waze to extract information from people using the navigation app and learn where congestion hot spots are. The city also partnered with Esri and developed a geospatial data visualization platform. The High Injury Network project, originating in San Francisco, maps a city's pedestrian and cyclist fatalities related to traffic incidents to identify risk factors and prevention strategies. Researcher Lisa Schweitzer of the USC Price School of Public Policy used Twitter in an award-winning paper analyzing stigma around public transit.

These advancements support agencies in improving traffic management and operations, and help the general public to better understand their local environment. More importantly, they inform evidence-based and data-driven decision-making in transportation policy and investment.

Social media has been leveraged in a myriad of studies for insights into travel demand estimation, mobility behavior assessment, traffic condition monitoring, and incidents and natural disasters modeling. However, only a few studies to date have used social media information for public transit analysis, mostly focusing on sentiment analysis to evaluate transit system performance from transit riders' perspectives.

PROOF OF CONCEPT AND NEXT STEPS

This project began as a proof of concept, as researchers sought to facilitate the use and integration of new, open transportation data for transit agencies in discovering and strengthening fundamental patterns of interactions between users and transit services. The ultimate goal is to provide a rich analytical platform to enable transit agencies to effectively explore insights from the integrated transportation data.

The researchers engaged two transit agencies, the Utah Transit Authority (UTA) and TriMet, to test the usability of the toolbox. Salt Lake City, Utah and Portland, Oregon were used as case studies in the platform for querying, navigating and exploring the interactions between transit users and services.

Leveraging machine learning and natural language processing techniques, the team retrieved Twitter data that are related to public transit systems and extracted sentence structures to geomap those tweets to their corresponding transit lines/stations. Combined with transit accessibility measures computed using GTFS, the tool enables us to identify the mismatch between the services the agency is providing versus what the transit users are experiencing.

Future goals for STAT's development include:

- Incorporating data from additional social media platforms like Facebook and Instagram,
- Including more transit agencies to reach other cities besides Portland and Salt Lake City,
- Adding more data sources, such as GTFS Realtime and GPS-based transit location trackers.

ABOUT THE AUTHORS

The research team consisted of Xiaoyue Cathy Liu (Principal Investigator), Qian Zuo and Shenruoyang Na, University of Utah; Ran Wei, University of California, Riverside; and Aaron Golub, Liming Wang and Jake Davis; Portland State University

ABOUT THE FUNDERS

This research was funded by the National Institute for Transportation and Communities, with additional support from Portland State University, the University of Utah, and community partners Assist Inc. and Unlimited Choices.

THE FULL REPORT and ONLINE RESOURCES

For more details about the study **Social-Transportation Analytic Toolbox (STAT) For Transit Networks**, download the full report at
https://nitc.trec.pdx.edu/research/project/1080

Use the STAT tool at:
http://xiaoyueliu.net:8002/

Photo by Transportation Research and Education Center at Portland State University

The National Institute for Transportation and Communities (NITC) is one of seven U.S. Department of Transportation national university transportation centers. NITC is a program of the Transportation Research and Education Center (TREC) at Portland State University. This PSU-led research partnership also includes the Oregon Institute of Technology, University of Arizona, University of Oregon, University of Texas at Arlington and University of Utah.