Semi-Supervised Machine Learning Framework for Fusing Georeferenced Data from Social Media and Community-Driven Applications

Christin Salley,[1] Neda Mohammadi, Ph.D., A.M.ASCE,[2] and John E. Taylor, Ph.D., M.ASCE[3]

[1] Ph.D. Student, School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA; e-mail: csalley3@gatech.edu
[2] City Infrastructure Analytics Director, School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA; e-mail: nedam@gatech.edu
[3] Frederick Law Olmsted Professor, School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA; e-mail: jet@gatech.edu

ABSTRACT

Digital forms of citizen communication with response organizations through social media are becoming more widespread during disasters. Public agencies often use this information to examine community discussions to assess, determine, and prioritize critical areas in need of assistance. However, limitations on harnessing high volumes of precise geolocation information from social media restricts their ability to locate and promptly delineate actionable insights. Here, we propose a semi-supervised machine learning framework that integrates Natural Language Processing (NLP) and spatiotemporal analytics to augment data from social media (Twitter) with a community-driven application (Waze) to achieve further evidence on location and type of emergency events. The framework is illustrated through a case study on Tropical Storm Zeta. This fusion provides increased context and may enhance the speed of emergency response. This study establishes a foundation for real-time crisis event detection, expanding organizations' response capacity in allocating resources and reducing harmful effects of disasters.