Social Media Operationalized for Locational Analytics: A Hands-on Tutorial Utilizing NLP, Topic Modeling, and GIS

Tutorial

Brian N. Hilton, Ph.D.
Claremont Graduate University
brian.hilton@cgu.edu

Anthony J. Corso, Ph.D.
California Baptist University
anthony.corso@calbaptist.edu

Sarah Osailan
Claremont Graduate University
sarah.osailan@cgu.edu

Extended Abstract

Social media such as Twitter, Reddit, and Facebook, have become de facto global communication channels to disseminate news, entertainment, and one’s self-revelations. However, though these organizations provide real-time data feeds that may be utilized in many ways, how does one actually go about using these data?

This tutorial will provide hands-on instruction regarding the collection, processing, and analysis of these data. First, we present an overview of our experiences using Natural Language Processing (NLP) techniques to operationalize social media data for use in locational analysis. Next, through a series of three exercises, we will demonstrate:

1) How the use of the open-source tool GATE may be customized for the NLP of a Tweet corpus;
2) How Topic Modeling tools such as Mallet can help facilitate an understanding of social behavior; and
3) How Geographic Information Systems (GIS) can enhance social media analysis via advanced locational analytics and reveal connections between these “conversations” and real-world events.

In addition to the tutorial activities, a concluding group activity and discussion will be facilitated focusing on the research areas revealed during the tutorial along with the key takeaways concerning the ongoing use of the techniques and tools utilized during the tutorial and the next learning steps for participants.

This tutorial is sponsored by the SIGGIS Mini-Track.

Keywords

Participants will need a laptop with Java installed. The only other requirement would be to have a browser installed and the ability to connect to the Internet.

Likely participants are faculty and graduate students in information systems, information science, social sciences / humanities, and other related disciplines, who have an interest in collecting and processing social media data using various NLP techniques, and then exploring and analyzing this enriched data with various locational analytic tools for research and teaching. Attendees can include participants with little knowledge of NLP or location analysis, ranging up to experts in those areas.