Androwis Abumoussa | University of Rochester, Rochester, NY`
ST12B-002: Forecasting Dynamic Group Behavior in Social Media ST12B
Department of Defense
Defense Advanced Research Projects Agency

Fount.in: Using geo-tagged social media to visual, monitor, learn, and predict group behavior
January 1, 2013 - May 2015

PI - Androwis Abumoussa - University of Rochester
co-PI - Henry Kautz, Ph.D. - University of Rochester

Project/Performance Site Primary Location
University of Rochester, Computer Studies Building
Department of Computer Science
734 Computer Studies Bldg. · P.O. Box 270226 · Rochester, NY 14627

Forecasting Dynamic Group Behavior in Social Media

Project Summary

Social media affords users with a low friction medium for communication. As information is broadcast between individuals social connections or graphs of social structures begin forming. Through these social graphs, analysis can be performed to understand how ideas, thoughts, diseases, and other communicable mediums are shared and ultimately, how they traverse a population. This proposal aims to understand the dynamics of idea, disease, or thought propagation occurs, focusing on identifying:

1) the origin of a communicable agent
2) the predicted path of growth or spread
3) the "health" of an individual with regard to a communicable agent

Prediction of group or individual behavior will be modeled as disease transmission. Real-time syndromic surveillance systems can be built and deployed on any scale of interest, from individuals to groups of individuals, to subsets of whole population.

By understanding the factors that affect these three points, the team at the University of Rochester aims to utilize the information provided by social media to classify the current 1) aggregate population behavior, 2) the behavior patterns of an individual, and then 3) the future state of interest for both population scale and the individual. The analysis of social media data allows us to filter both the conversations and interactions that occur on social media with regard to geography.

The current state of individuals:
The first problem is identifying communicable agents. We do this by analysing social graph structures for high frequency occurrences. By applying statistical analysis to identify communicable agents, systems are then able to analyze over a million features that every tweet, facebook status, or foursquare checkin contains to predict the "state" (whether they're infected or not) of individuals. The system can handle passively

observed data, as well as user provided feedback to update the statistic models.

After a communicable agent is identified, the system is then able to correlate this with measured actions or saturation points within social structures. As the system observes whether a communicable agent will spread, to whom, where, and why, we're then able to predict the future behavior of an indivdual.

Research has shown that Fount.in is able to predict when healthy people were about to fall ill - and then tweet about it - with about 90 per cent accuracy out to eight days in the future. Fount.in is unique tool designed to crowd source the monitoring of diseases in realtime. Preliminary research has shown that these techniques can identify infectious individuals with over 90% accuracy and builds the framework for a distributed and effective pre-processing tool for syndrome classification.

Intellectual Merit:
Much of this research focuses on both the volume of the information as well as the free-text format by which most of the data exists. To address the volume issue, the system will provide an efficient mechanism for machine learning algorithms to work in unsupervised fashions on free text and colloquial vernacular.

Secondly the system will couple sparse human input through labeling with machine learning. An interface by which we can crowd source the learning of features and then allow machines to act on the volumes. We can train multiple types of AI to process this information so neural networks, SVM's etc can be applied.

Broader Impact:
We will demonstrate how computers analyze and then present social media data to empower the community and policy makers to predict when points of inflections occur. Fount.in is a unique platfrom designed specifically to understand interactions within social media to improve our ability to predict group dynamics. By examining geo-tagged social media interactions the team at the University of Rochester have created an epidemic surveillance system which healthcare providers and public policy officials can use to inform their decisions, in real-time and we hope to generalize this.

Our artificial intelligence models observe and analyze patterns of behavior over free-text, social, and geospatial interactions to build a health model for individuals, generating an 'artificial differential diagnosis'. These models are then made available for a variety of primary clinical and secondary uses within government organizations and business entities, including syndromic surveillance and communicable agent refinement.

Using big data platforms and user generated content in lieu of constant monitoring or surveying by intelligence agencies, we are now able to understand how communicable agents spread from person to person and the implications this has on a population's behavior as a whole. Using geo-tagged social media content we're able to do this in real-time and with flexible and modular models so that the underlying artificial intelligence can be adapted to different social vernaculars and health care settings.