[MS Thesis Defense]

Computational Social Analysis: Social Unrest Prediction Using Textual Analysis of News

Ehsan Alikhani
MS candidate, Systems Science
Binghamton University

Monday May 5th, 2014   3:00-4:00pm
Engineering Building R3 (SSIE Conference Room)

Revolutions, rebellions and upheavals not only cost human life and cause misery, but also strongly affect the stability of nations, make investments insecure and intensify the cycle of poverty. Social NGOs and responsible governments are interested in predicting social unrest to lessen or even avoid its negative consequences, and to control the spread of violence. The goal of this research is to predict the risk of social unrest in a selected set of countries by analyzing tone in media news articles. For several years, research on social unrest has been mostly descriptive with limited case studies. In recent years, a few pioneering studies have started to use data mining approaches on large scale data sets to discover patterns in the emergence of social unrest. Despite relatively accurate predictions in some cases, they rely on limited number of factors like food price, GDP and population to create prediction models. In this research, an extensive data mining work has been done on a newly published data base named GDELT which contains about 250 million analyzed news articles of social events since 1979. Each article is converted to a record containing the two actors involved in an event, the tone of the article describing it and geographic attributes of the event. The prediction model developed in this study is based upon a simple idea: If the tone of the articles describing major role players inside a country or foreign interactions of a country is changing, it might affect the stability of the nation. The role player can be military, civilians, and international entity or a local organization. For the first step in forming the model, influential social role players in each society were discovered and their effects on social unrest were analyzed. Next the trend of the tones and frequencies of news stories about each role player were identified. Finally, discriminant analysis and multivariate regression were conducted to discover dynamic patterns of interaction. The prediction result showed that the developed model was able to predict major social unrests with 70~80% accuracy for certain cases within a period of three months into the future.

For more information, contact Hiroki Sayama (sayama@binghamton.edu)
http://coco.binghamton.edu/