Indiana University-Purdue University Indianapolis

Department of Mathematical Sciences

STATISTICS SEMINAR

12:15pm—1:15pm, Tuesday, November 12, 2019
LD 018

Speaker: **Sunandan Chakraborty, Assistant Professor of Data Science**
School of Informatics and Computing, IUPUI

Title: **Recurrent Event Network: A News Event-driven Predictive Models For Forecasting Socio-economic Indicators**

Abstract:

Accurately predicting socio-economic indicators, such as crop prices and disease incidence, is critical for effective policy making. It is however challenging because the indicators are highly volatile and have complex dependence on real-world events. In this paper, we address this challenge by leveraging corpora of news articles, which report several of the events that affect the socio-economic indicators. To capture these events in a meaningful way, we propose an event-based generative model. Unlike general purpose generative models, such as topic models, our generative model exploits the unique discourse structure of news articles, which report a central event in the headline and/or the lead paragraph. Our model assumes that the central event of each article is sampled from a distribution over all possible event classes, which are abstract groupings of similar events. These event classes manifest in news articles as event triggers: specific words (or phrases) describing the actions in the event. We use the extracted events to generate time-stamped event embeddings—low-dimensional vector representations of events occurring in a news corpus. Then, we introduce the concept of Recurrent Event Networks (RENs), which uses re-designed recurrent neural network (RNN) architectures and event embeddings to predict a time-series of a socio-economic indicator of interest. We trained RENs across two broad classes of socio-economic indicators—crop prices and disease incidence levels—on a large corpus from a major Indian news source covering 10 years of data. Our results show that RENs outperform a suite of benchmark methods. Our event embeddings are instrumental in achieving this performance because other general-purpose embeddings had much worse performance. In our experiments, we found that RENs with event embeddings had 48.77%, 34.06%, and 53.60% lower root mean-square errors

(RMSE) than RENs with word2vec, LDA, and nonnegative matrix factorization (NMF) embeddings, respectively, for crop price predictions. Similar results were observed for disease incidence predictions. RENs also significantly outperformed linear predictive models.