| TOPIC | Data Science & Analytics for Intrusion Detection |
|---|---|
| ORGANIZERS | Student Leadership Council and Faculty of ACIT Institute and TECHLAV Center |
| AREA | Machine Learning, Data Mining, Time series Prediction |
| SPEAKER | Norbert Agana |
| DATE | Friday June 16, 2017 |
| TIME | 3:00 – 4:00 P.M. (EST) |
| VENUE | Fort IRC 410, North Carolina A&T State University |
| FEES | No Charge |

## SYNOPSIS

Time Series Prediction plays a significant role in almost all fields of science and engineering including climate and meteorology, financial and business, industrial applications, reliability forecasting, etc. The underlying dynamics and time series data that generates the processes of these systems are generally complex. Traditional techniques for time series prediction are based on an implicit assumption of a stationarity observed process, which is a contradiction to most real processes that are often chaotic and nonstationary. Artificial neural networks with backpropagation learning algorithms have shown great promise over the last two decades in modeling nonlinear time series. However, the issue of nonconvex optimization ensues when two or more hidden layers are required for highly complex phenomena. To learn the complex features of these systems, more hidden layers are usually added to the network, but this is not often successful due to problems encountered during the training process. This research addresses this issue by developing a novel deep-learning-based time series prediction model inspired by recent advancement in deep learning training techniques. An unsupervised greedy layer-wise training methodology is employed to train the deep networks. Deep learning algorithms trained using this approach have shown empirically to avoid getting stuck in local solutions and hence, can improve prediction accuracy. We also showed that series decomposition, when integrated with these techniques, results in improvement of model accuracy. This research also shows how this algorithm can be applied to climate data. Two real-world datasets are considered for empirical evaluation of the model. First, the model is applied to forecast a long-term drought using a standardized streamflow index (SSI) as the climate variable. The second application characterizes the solar activity as measured by the annual mean value of the sunspot number (Wolf number). The second data set is often used in the literature as a benchmark for testing new statistical techniques.

## ABOUT THE SPEAKER

Norbert Agana is a Ph.D. student at North Carolina A&T State University. He is currently a graduate research assistant at the Autonomous Control and Information Technology (ACIT) Institute. His research interest is in Machine Learning, Data Mining and Time Series Analysis and their applications to extreme events prediction. He received his BSc in Mathematics and MSc in Electrical Engineering from University of Cape Coast, Ghana and Tuskegee University, Alabama respectively.