COLLEGE OF ENGINEERING AND COMPUTER SCIENCE
FLORIDA ATLANTIC UNIVERSITY

Announces the Ph.D. Dissertation Defense of

Eric Golinko

for the degree of Doctor of Philosophy (Ph.D.)

"Generalized Feature Embedding Learning for Clustering and Classification"

June 22, 2018, 12 p.m.
Engineering East, EE 405
777 Glades Road
Boca Raton, FL

DEPARTMENT
Computer & Electrical Engineering and Computer Science

CHAIR OF THE CANDIDATE'S PH.D. COMMITTEE
Xingquan Zhu, Ph.D.

PH.D. SUPERVISORY COMMITTEE
Ankur Agarwal, Ph.D.
Mehrdad Nojoumian, Ph.D.
Shihong Huang, Ph.D.
Dingding Wang, Ph.D.

ABSTRACT OF DISSERTATION

Generalized Feature Embedding Learning for Clustering and Classification

Data comes in many different shapes and sizes. This may include, numerical, categorical, and text. In order to be able to model this data with machine learning algorithms, it is required that the data is typically in numeric form. Therefore, for data that is not originally numerical, it must be transformed to be able to be used as input into these algorithms. Along with this transformation, it is often desirable to reduce the number of features that are being trained in a model to eliminate noise and reduce time in training. This problem of high dimensionality can be approached through feature selection, feature extraction, or feature embedding. This dissertation focuses on feature extraction and embedding which are techniques that utilize a mathematical transformation of the data into a represented space. As a byproduct of using a new representation, we are able to reduce the dimension greatly without sacrificing performance. We are motivated to illustrate a methodology that can be applied to any data type with little pre-processing. The methods we develop can be applied in unsupervised, supervised, incremental, and deep learning contexts. Using 28 benchmark datasets we are able to show the efficacy of our method when compared to baseline. The techniques we develop contribute to the field of dimension reduction and feature embedding. This is accomplished by three main vital components. The first being a class partitioned row and feature product representation of one-hot encoded data. Secondarily, the derivation of a weighted adjacency matrix based on class label relationships. Finally, by the inner product of these aforementioned values, we are able to condition the one-hot encoded data generated from the original data prior to eigenvector decomposition. The use of class partitioning and adjacency enable subsequent projections of the data to be trained more effectively when compared side-to-side to baseline algorithm performance.

BIOGRAPHICAL SKETCH
Born in Brooklyn, New York
B.S., Florida Atlantic University, Boca Raton, Florida, 2009
M.S., Florida Atlantic University, Boca Raton, Florida, 2012
Ph.D., Florida Atlantic University, Boca Raton, Florida, 2018

CONCERNING PERIOD OF PREPARATION & QUALIFYING EXAMINATION

Time in Preparation: 2016 - 2018

Qualifying Examination Passed: Fall 2015

Published Papers:

Eric Golinko and Xingquan Zhu. "Generalized Feature Embedding for Supervised, Unsupervised, and Online Learning Tasks." *Information Systems Frontiers* (2018): 1-18. (Impact Factor: 2.521).

Eric Golinko and Xingquan Zhu. "GFEL: Generalized feature embedding learning using weighted instance matching." *2017 IEEE International Conference on Information Reuse and Integration (IRI)*. San Diego, CA, August 4-7, 2017.

Eric Golinko, Thomas Sonderman, and Xingquan Zhu. "CNFL: Categorical to Numerical Feature Learning for Clustering and Classification." *2017 IEEE Second International Conference on Data Science in Cyberspace (DSC), Shenzhen, China, June 26-29,* 2017.