



**COLLEGE OF ENGINEERING
AND COMPUTER SCIENCE**
FLORIDA ATLANTIC UNIVERSITY

Announces the Ph.D. Dissertation Defense of

Joseph Daniel Prusa

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

"Enhancement of Deep Neural Networks and Their Application to Text Mining"

March 26, 2018, 11 a.m.
Engineering East, Room 405
777 Glades Road
Boca Raton, FL

DEPARTMENT:

Computer and Electrical Engineering and Computer Science

CHAIR OF THE CANDIDATE'S PH.D. COMMITTEE:

Taghi M. Khoshgoftaar, Ph.D.

PH.D. SUPERVISORY COMMITTEE:

Mehrdad Nojournian, Ph.D.

Dingding Wang, Ph.D.

Xingquan Zhu, Ph.D.

ABSTRACT OF DISSERTATION

Enhancement of Deep Neural Networks and Their Application to Text Mining

Many current application domains of machine learning and artificial intelligence involve knowledge discovery from text, such as sentiment analysis, document ontology and spam detection. Humans have years of experience and training with language, enabling them to understand complicated, nuanced text passages with relative ease. A text classifier attempts to emulate or replicate this knowledge so that computers can discriminate between concepts encountered in text. Learning high-level concepts from text, such as those found in many applications of text classification, is a challenging task due to the many challenges associated with text mining and classification. Recently, classifiers trained using artificial neural networks have been shown to be effective for a variety of text mining tasks. Convolutional neural networks have been trained to classify text from character-level input, automatically learn high-level abstract representations and avoiding the need for human engineered features. This dissertation proposes two new techniques for character-level learning, log(m) character embedding and convolutional window classification. Log(m) embedding is a new character-vector representation for text data that is more compact and memory efficient than previous embedding vectors. Convolutional window classification is a technique for classifying long documents, i.e. documents with lengths exceeding the input dimension of the neural network. Additionally, we investigate the performance of convolutional neural networks combined with long short-term memory networks, explore how document length impacts classification performance and compare performance of neural networks against non-neural network-based learners in text classification tasks.

BIOGRAPHICAL SKETCH

B.S. University of Florida, Gainesville, Florida, 2012

M.S. Florida Atlantic University, Boca Raton, Florida, 2015

Ph.D. Florida Atlantic University, Boca Raton, Florida, 2018

CONCERNING PERIOD OF PREPARATION & QUALIFYING EXAMINATION

Time in Preparation: 2015 - 2018

Qualifying Examination Passed: Fall 2015

Published Papers:

Prusa, Joseph D., and Taghi M. Khoshgoftaar. "Training Convolutional Networks on Truncated Text." In Tools with Artificial Intelligence (ICTAI), 2017 IEEE International Conference on, pp. 330-335. 2017.

Prusa, Joseph D., and Taghi M. Khoshgoftaar. "Improving Deep Neural Network Design With new Text Data Representations." Journal of Big Data, vol. 4:7, 2017.

Prusa, Joseph D., and Taghi M. Khoshgoftaar. "Deep Neural Network Architecture for Character-Level Learning on Short Text." In 30th Florida Artificial Intelligence Research Society (FLAIRS), International Conference, pp. 353-358. 2017.

Prusa, Joseph D., and Taghi M. Khoshgoftaar. "Designing a Better Data Representation for Deep Neural Networks and Text Classification." In Information Reuse and Integration (IRI), 2016 IEEE 17th International Conference on, pp. 411-416. 2016.