



Announces the Ph.D. Dissertation Defense of

## Sahar Sohangir

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

### “Machine Learning Methods to Understand Textual Data”

Sept. 25, 2018, 10 a.m.

777 Glades Rd., Engineering East (EE), 405  
Boca Raton, FL

DEPARTMENT: Computer and Electrical Engineering  
& Computer Science

ADVISOR: Dingding Wang, Ph.D.

PH.D. SUPERVISORY COMMITTEE:

Dingding Wang, Ph.D., Chair  
Taghi M. Khoshgoftaar, Ph.D.  
Xingquan Zhu, Ph.D.  
Shihong Huang, Ph.D.

ABSTRACT OF DISSERTATION

#### Machine Learning Methods to Understand Textual Data

Artificial intelligence scientists are always trying to figure out how machines can be developed to understand human language as it is spoken. Natural language processing (NLP) is an area in artificial intelligence that investigates how to program computers to process and analyze large amounts of textual data. For instance, a human can easily understand how similar two documents are, but how can a machine measure the similarity between two documents? Cosine similarity is one of the most commonly use similarity measurements in NLP, however it has some difficulties in high dimensional data. This dissertation proposes a new similarity measurement that can alleviate the problem of cosine similarity in high dimensional data. Similarity measurement plays a significant rule in natural language processing and is broadly used in information retrieval, text classification, document clustering, etc.

In recent years the text mining field has gained a great deal of attention due to the availability of massive amounts of data in a variety of forms such as social networks, patient records, healthcare insurance data, news outlets, etc. This volume of text contains an incredible source of information and knowledge. As a result, there is a desirable need to design methods and algorithms to effectively process this massive amount of text in a wide variety of applications. Processing of this tremendous volume of mostly unstructured data is not a straightforward function. Recently, deep learning approaches have emerged as a powerful tool in analyzing textual data due to the advantages they provide over other methods. One of these advantages is that features are learned hierarchically during the process of deep learning instead of the feature engineering that is required in data mining. Additionally, in deep learning methods, each word is considered as part of a

sentence. In this way, relevant information contained in word order, proximity, and relationships is not lost. For instance, the application of neural network techniques on social networking websites can reveal a significant amount of information. The modern stock market is an example of these social networks. They are a popular place to increase wealth and generate income, but the fundamental problem of when to buy or sell shares, or which stocks to buy has not been solved. It is very common among investors to have professional financial advisors, but what is the best resource to support the decisions these people make? This dissertation investigates if deep learning methods such as convolutional neural network, long short-term memory, and doc2vec can answer these questions.

#### BIOGRAPHICAL SKETCH

Born in Tehran, Iran  
B.S., Azad University, Tehran, Iran, 2006  
M.S., Azad University, Tehran, Iran, 2010  
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:

Sahar Sohangir, Dingding Wang, Anna Pomeranets, and Taghi M. Khoshgoftaar. Big data: deep learning for financial sentiment analysis. *Journal of Big Data*, 5(1): 3, 2018.

Sahar Sohangir and Dingding Wang. Document understanding using improved sqrt-cosine similarity. In *Semantic Computing (ICSC)*, 2017 IEEE 11th International Conference on, pages 278–279. IEEE, 2017.

Sahar Sohangir, Nicholas Petty, and Dingding Wang. Financial sentiment lexicon analysis. In *Semantic Computing (ICSC)*, 2018 IEEE 12th International Conference on, pages 286–289. IEEE, 2018.

Sahar Sohangir and Dingding Wang. Finding expert authors in financial forum using deep learning methods. In *2018 Second IEEE International Conference on Robotic Computing (IRC)*, pages 399–402. IEEE, 2018.

Sahar Sohangir and Dingding Wang. Finding expert authors in financial forum using deep learning methods. In *2018 Second IEEE International Conference on Robotic Computing (IRC)*, pages 399–402. IEEE, 2018.