Tom Hartvigsen

twhartvigsen@wpi.edu thartvigsen.github.io

Interests: Deep Learning, Recurrent Neural Nets, Interpretability, Time Series, Reinforcement Learning.

Education

Worcester Polytechnic Institute, *Worcester, MA*
Ph.D., Data Science 2021
Advisors: Elke Rundensteiner, Xiangnan Kong

SUNY Geneseo, *Geneseo, NY*
B.A., Applied Mathematics 2016
BioMathematics minor
Advisors: Chris Leary, Kirk Anne

Experience

Graduate Research Fellow, *Worcester Polytechnic Institute* Aug 2016 - 2021
Studying and developing recurrent models for a variety of challenging sequence classification tasks.
Supervisors: Prof. Elke Rundensteiner, Prof. Xiangnan Kong

Research Intern (Machine Learning - NLP), *UMass Medical School* Sep 2018 - Aug 2019
Worked on auto-summarization of clinical trial eligibility criteria for recommendation in new trials.
Supervisor: Dr. Jomol Matthew

NSF REU Intern, *University of Arizona, Department of Environmental Science* Summer 2015
Built an instance-segmentation model for remotely-captured images of creosote bushes to model the effects of drought over time in arid environments.
Supervisor: Prof. Shirley Papuga

Research Assistant, *SUNY Geneseo* 2014 - 2016
Modeled infection spread on graphs, built a graph dataset from IMDB, mined song lyrics for text features useful for discriminating genres and artists.
Supervisors: Prof. Chris Leary, Dr. Kirk Anne

Publications

In-submission

1. *Learning to Selectively Update State Neurons in Recurrent Networks.*
Thomas Hartvigsen, Cansu Sen, Xiangnan Kong, Elke Rundensteiner.
2. *Deep Biased Positive Unlabeled Learning of Sequential Data.*
Walter Gerych, **Thomas Hartvigsen**, Luke Buquicchio, Elke Rundensteiner.
3. *Explainable Document Classification with Human-guided Attention.*
Cansu Sen, **Thomas Hartvigsen**, Xiangnan Kong, Elke Rundensteiner.
4. *Learning Similarity-Preserving Word Meta-Embedding.*
Jidapa Thadajarassiri, Cansu Sen, **Thomas Hartvigsen**, Xiangnan Kong, Elke Rundensteiner.

Peer-Reviewed Publications

1. *Recurrent Halting Chain for Early Multi-label Classification.*
Thomas Hartvigsen, Cansu Sen, Xiangnan Kong, Elke Rundensteiner.
ACM SIGKDD Conference on Knowledge Discovery and Data Mining (**KDD**), 2020.

2. *Human Attention Maps for Text Classification: Do Humans and Neural Networks Focus on the Same Words?*
   Cansu Sen, **Thomas Hartvigsen**, Biao Yin, Xiangnan Kong, Elke Rundensteiner.
   Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (**ACL**), 2020.

3. *Clinical Performance Evaluation of a Machine Learning System for Predicting Hospital-Acquired Clostridium Difficile Infection.*
   Erin Teeple, **Thomas Hartvigsen**, Cansu Sen, Kajal Claypool, Elke Rundensteiner.
   International Conference on Health Informatics (**HEALTHINF**), 2020. **Best poster award**.

4. *Adaptive-Halting Policy Network for Early Classification.*
   Thomas Hartvigsen, Cansu Sen, Xiangnan Kong, Elke Rundensteiner.
   ACM SIGKDD Conference on Knowledge Discovery and Data Mining (**KDD**), 2019.

5. *Patient-Level Classification of Clinical Note Sequences Guided by Attributed Hierarchical Attention.*
   Cansu Sen, **Thomas Hartvigsen**, Xiangnan Kong, Elke Rundensteiner.
   IEEE International Conference on Big Data (**BigData**), 2019.

6. *Learning Temporal Relevance in Longitudinal Medical Notes.*
   Cansu Sen, **Thomas Hartvigsen**, Xiangnan Kong, Elke Rundensteiner.
   IEEE International Conference on Big Data (**BigData**), 2019.

7. *Comparing General and Locally-Learned Word Embeddings for Clinical Text Mining.*
   Jidapa Thadajarassiri, Cansu Sen, **Thomas Hartvigsen**, Xiangnan Kong, Elke Rundensteiner.
   IEEE International Conference on Biomedical and Health Informatics (**BHI**), 2019.

8. *Early Diagnosis Prediction with Recurrent Neural Networks.*
   Daniel Johnston*, Liubou Klindziuk*, Lolita Nazarov*, **Thomas Hartvigsen**, Elke Rundensteiner.
   IEEE MIT Undergraduate Research Technology Conference (**URTC**), 2019. **Best paper runner up**.

9. *Detecting MRSA Infections by Fusing Structured and Unstructured Electronic Health Record Data.*
   Thomas Hartvigsen, Cansu Sen, Elke Rundensteiner.
   Communications in Computer and Information Science (**CCIS**) 1024, 2018.

10. *Early Prediction of MRSA Infections using Electronic Health Records.*
    Thomas Hartvigsen, Cansu Sen, Sarah Brownell*, Erin Teeple, Xiangnan Kong, Elke Rundensteiner.
    International Conference on Health Informatics (**HEALTHINF**), 2018. Short-listed for Best Student Paper.

11. *Handling Missing Values in Multivariate Time Series Classification.*
    Julia Friend*, Alec Hauck*, Sruthi Kurada*, Cansu Sen, **Thomas Hartvigsen**, Elke Rundensteiner.
    IEEE MIT Undergraduate Research Technology Conference (**URTC**), 2018.

12. *CREST - Risk Prediction for Clostridium Difficile Infection Using Multimodal Data Mining.*
    Cansu Sen, **Thomas Hartvigsen**, Kajal Claypool, Elke Rundensteiner.
    European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (**ECML**), 2017.

*Undergraduate collaborator.

Honors and Awards

Best Poster, HEALTHINF	2020
Graduate Student Travel Grant, WPI	2020
IMA Travel Grant, University of Minnesota	2019
KDD Student Travel Grant, NSF and ACM	2019
Graduate Student Travel Grant, WPI	2019
Best Poster, Graduate Research Innovation and Exchange, WPI	2019
People's Choice Poster Award, Graduate Research Innovation and Exchange, WPI	2018
Graduate Student Travel Grant, WPI	2018
People's Choice Poster Award, Graduate Research Innovation and Exchange, WPI	2017

Graduate Student Travel Grant, WPI 2017
GAANN Ph.D. Fellowship, U.S. Department of Education 2016-2021

Invited Talks

Adaptive-Halting Policy Networks for Early Classification
Florida State University, Panama FL 2020
MITRE, Bedford MA 2020
University of Minnesota, Minneapolis MN 2019
New England Machine Learning Day Poster, Northeastern University, Boston MA 2019

Recurrent Models for Clinical Time Series 2019
Lightning Talk at WPI Arts & Sciences Week

Selective Activation in Recurrent Neural Networks 2019
Data Science Research Group Colloquium, WPI

Early Prediction of MRSA Infections Using Electronic Health Records 2018
HEALTHINF conference, Funchal, Madeira Island, Portugal

Other Talks

Partial Recurrent State Updates for Irregular Multivariate Time Series 2019
Graduate Research Innovation and Exchange Poster Session, WPI

Introduction to PyTorch and Deep Learning 2019
PyTorch tutorial given to NSF-funded summer REU students with Walter Gerych, WPI

JokeR: A Recurrent Joke Generator 2018
Deep Learning final project with Thanh Tran and Sanket Gujar, WPI

Adaptively-Halting RNN for Tunable Earliness in Multivariate Time Series Classification 2018
Graduate Research Innovation and Exchange Poster Session, WPI

CREST - Risk Prediction for Clostridium Difficile Infection Using Multimodal Data Mining 2017
Graduate Research Innovation and Exchange Poster Session, WPI

Teaching

NSF REU Project Advisor, WPI. Summers of 2017-19
Students: L. Klindziuk, D. Johnston, L. Nazarov, J. Friend, A. Hauck, S. Kurada, S. Brownell, S. Tocci.
Outcomes: One paper per summer from 2017, 2018, and 2019.
Teaching Assistant, SUNY Geneseo, Modeling Biological Systems (2x) and BioStats (1x). 2015-2016
Modeling Biological Systems, SUNY Geneseo 2016
Guest lecturer: taught Percolation Models, including an in-class exercise in R.

Service

Graduate Student Council of Arts & Sciences, WPI 2018-2020
Data Science Representative on the Graduate Student Government Senate, WPI 2018
Data Science Graduate Student Council, WPI 2016-2019

Technical Skills

Programming: Python, R, LaTeX, SQL.
Frameworks: PyTorch, TensorFlow, Scikit-learn, NumPy.