Tom Hartvigsen

twhartvigsen@wpi.edu thartvigsen.github.io

Interests: Machine Learning, Recurrent Neural Nets, NLP, Conditional Computation, Sustainability.

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

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

Machine Learning Research Intern, *University of Massachusetts Medical School* 2018 - 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: Dr. 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: Dr. Chris Leary, Dr. Kirk Anne

Publications

In-submission

1. *Reducing Computation in Recurrent Networks by Selectively Updating State Dimensions.*
Thomas Hartvigsen, Cansu Sen, Xiangnan Kong, Elke Rundensteiner.
International Conference on Machine Learning (**ICML**), 2020.

2. *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.

3. *Deep Biased Positive Unlabeled Learning of Sequential Data.*
Walter Gerych, **Thomas Hartvigsen**, Luke Buquicchio, Elke Rundensteiner.
ACM SIGKDD Conference on Knowledge Discovery and Data Mining (**KDD**), 2020.

4. *Explainable Document Classification with Human-guided Attention.*
Cansu Sen, **Thomas Hartvigsen**, Xiangnan Kong, Elke Rundensteiner.
ACM SIGKDD Conference on Knowledge Discovery and Data Mining (**KDD**), 2020.

5. *Similarity-Preserving Meta-Embedding.*
Jidapa Thadajarassiri, Cansu Sen, **Thomas Hartvigsen**, Xiangnan Kong, Elke Rundensteiner.

Peer-Reviewed Publications

1. *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.
   Annual Conference of the Association for Computational Linguistics (**ACL**), 2020.

2. *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**.

3. *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.

4. *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.

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

6. *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.

7. *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**.

8. *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.

9. *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.

10. *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.

11. *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.

*Student under my supervision.

Selected Awards

Best Poster Award, HEALTHINF	2020
Graduate Student Travel Grant, WPI	2020
IMA Travel Grant, Institute for Mathematics and its Applications, U. of Minn.	2019
KDD 2019 Student Travel Grant, NSF and ACM	2019
Graduate Student Travel Grant, WPI	2019
Best Poster Award, 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

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.
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.

Technical Skills

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