

Dustin Wright

Curriculum Vitae

PERSONAL DETAILS

Phone +45 93 50 91 78
Mail dw@di.ku.dk
GitHub <https://github.com/dwright37>
Website dustinbwright.com

PUBLICATIONS

Wright, D., & Augenstein, I. (2020). Semi-Supervised Exaggeration Detection of Health Science Press Releases. In *EMNLP 2021*. Association for Computational Linguistics.

Wright, D., & Augenstein, I. (2020). CiteWorth: Cite-Worthiness Detection for Improved Scientific Document Understanding. In *Findings of ACL 2021*. Association for Computational Linguistics.

Holm, A., **Wright, D.**, & Augenstein, I. (2020). Longitudinal citation prediction using temporal graph neural networks. arXiv preprint arXiv:2012.05742

Wright, D., & Augenstein, I. (2020). Transformer based multi-source domain adaptation. In *EMNLP 2020*. Association for Computational Linguistics.

Atanasova, P.*, **Wright, D.***, & Augenstein, I. (2020). Generating label cohesive and well-formed adversarial claims. In *EMNLP 2020*. Association for Computational Linguistics.

* denotes equal contribution

Wright, D., & Augenstein, I. (2020). Claim check-worthiness detection as positive unlabelled learning. In *Findings of EMNLP*. Association for Computational Linguistics.

Badal, V. D., **Wright, D.**, Katsis, Y., Kim, H. C., Swafford, A. D., Knight, R., & Hsu, C. N. (2019). Challenges in the construction of knowledge bases for human microbiome-disease associations. *Microbiome*, 7(1), 1-15.

Wright, D., Katsis, Y., Mehta, R., & Hsu, C. (2019). Normco: Deep disease normalization for biomedical knowledge base construction. AKBC 2019.
Best Application Paper

Koh, E.S., Dubnov, S., **Wright, D.** (2018). Rethinking recurrent latent variable model for music composition. *CoRR*, abs/1810.03226.

Bhide, A., **Wright, D.**, & Ozturk, Y. (2016). Per-packet rate adaptation for wireless video. *Signal, Image and Video Processing*, 10(7), 1273-1278.

Wright, D., Yan, X., Srinivas, P., Kashani, A., & Ozturk, Y. (2015). A cloud to mobile application for consumer behavior modification. *Procedia Computer Science*, 62,

343-351.

Yan, X., **Wright, D.**, Kumar, S., Lee, G., & Ozturk, Y. (2015). Real-time residential time-of-use pricing: a closed-loop consumers feedback approach. In *Green Technologies Conference (GreenTech), 2015 Seventh Annual IEEE* (pp. 132-138). IEEE.

Yan, X., **Wright, D.**, Kumar, S., Lee, G., & Ozturk, Y. (2015). Enabling consumer behavior modification through real time energy pricing. In *Pervasive Computing and Communication Workshops (PerCom Workshops), 2015 IEEE International Conference on* (pp. 311-316). IEEE.

Massai, S., Routhu, S., **Wright, D.**, Moon, K. S., Ozturk, Y., & Lee, S. Q. (2015). A Wireless Visual Attention Brain Signal Monitoring System. In *MATEC Web of Conferences* (Vol. 32, p. 04005). EDP Sciences.

EDUCATION

PhD. Computer Science

Oct 2022

University of Copenhagen

Researching natural language understanding and machine learning, focusing on fact checking, scientific language understanding, and domain adaptation. Advisor: Isabelle Augenstein.

MSc. Computer Science

June 2019

University of California, San Diego (GPA: 3.92)

Specialization in natural language processing, performing research on automatic biomedical knowledge base construction.

BSc. Computer Engineering

Dec 2014

San Diego State University (GPA: 3.97)

Performed research on applied machine learning. Investigated adaptive energy pricing and models for brain computer interfaces using EEG.

RESEARCH PROJECTS

Machine Understanding of Scientific Language

Oct 2019 - Present

University of Copenhagen

PhD project focusing on machine understanding of scientific text for fact checking. To date the work has focused on developing methods for check-worthiness detection, domain adaptation, generating difficult adversarial claims for fact checking, building resources for scientific fact checking, and exaggeration detection in science communication.

Generating Scientific Claims for Automated Fact Checking

Jun 2021 - Nov 2021

Allen Institute for AI

Project focused on automatically generating scientific claims for improved scientific fact checking. We developed a novel method of generating negations of scientific claims using a knowledge base, and demonstrated that we could achieve zero-shot fact checking while generating high quality claims.

Biomedical Knowledge Base Construction using Active Learning

Jun 2019 - Sep 2019

IBM Research

We devised a method to quickly construct highly curated datasets to enable biomedical knowledge base construction. The method, named *BioAct*, is based on a partnership between automatic annotation methods and subject matter experts and uses active learning to

create training datasets in the biological domain. We show that *BioAct* is useful for quickly constructing high quality biomedical knowledge bases, evaluating our method on a knowledge base construction task.

NormCo: Deep Disease Normalization

Jan 2018 - May 2019

UC San Diego

Designed and implemented a deep learning model for resolving disease mentions in text to unique concept IDs. The architecture made use of dense word embeddings and a recurrent neural network to model semantic information and ensure a coherent set of diseases was predicted within a document. Our model achieved state of the art results on two disease normalization corpora, and was accepted at AKBC 2019.

INVITED TALKS

Cite-Worthiness Detection for Improved Scientific Document Understanding (ETH Zürich, March 1, 2021)

NormCo: Deep Disease Normalization for Biomedical Knowledge Base Construction (10/07/2019, IBM Research)

WORK EXPERIENCE

PhD Researcher

Oct 2019 - Present

University of Copenhagen

- Working with Prof. Isabelle Augenstein on scientific language understanding for fact checking.
- 5 publications improving state of the art for check-worthiness detection, generating adversarial claims, understanding multi-source domain adaptation with pretrained transformers, cite-worthiness detection, and scientific exaggeration detection

Research Intern

Summer 2021

Allen Institute for AI

- Built a model for automatically generating scientific claims based on BART
- Developed method for generating negations of scientific claims which outperformed strong baselines on negation generation by a large margin
- Demonstrated that a model trained on our generated claims achieves within 90% of the performance of a model trained on in-domain manually written claims

Research Intern

Summer 2019

IBM Research

- Used active learning to build a dataset of several thousand annotations of antimicrobial drug resistance facts.
- Showed that we can up to double the amount of labels in a dataset using active learning in 75% of the time using less than 10% of the number of annotators.
- Demonstrated that labels acquired through active learning still improve a model's ability to perform its downstream task (knowledge base completion).

Graduate Student Researcher

2017-2019

UCSD Center for Microbiome Innovation

- Worked with a team of 4 to populate a knowledge base of disease-bacteria associations mined from millions of PubMed abstracts.

- Designed and implemented a deep learning based entity normalization algorithm using PyTorch which outperforms state of the art methods on benchmark datasets by 4.1%.
- Awarded Best Application Paper at AKBC 2019 conference.

Graduate Teaching Assistant

2017-2017

UCSD CSE Department

- Teaching assistant for CSE250A: Probabilistic Reasoning and Decision-Making
- Held regular office hours and assisted students in learning the material for a graduate AI course.
- Assisted students with the following topics: belief networks, linear and logistic regression, expectation maximization (EM), hidden markov models, reinforcement learning

Software Engineer II

2015-2017

BAE Systems Inc.

- Designed and implemented a web based chat application using React.js and Node.js which was used by first responders to help coordinate their efforts.
- Built several back end data processing features, including a pipeline which transformed various image formats into geo-located PDFs.

Research Assistant

2014-2015

SDSU Research Foundation

- Created a novel video chat system which adapted the video size, frame rate, and bandwidth based on the capacity of the communication channel. The system learned about user preferences in regards to scaling using support vector machines and made scaling decisions based on a learned model of user behavior.
- Worked with a team of 6 to design and implement a cloud and mobile based adaptive energy pricing system. Energy prices were forecast using support vector regression and users could view their current price and predicted price based on their energy usage.