Seojin Bang

860-709-4109 | seojinb@alumni.cmu.edu | seojinb.com

PROFILE

Broad research experience in machine learning and deep learning. Solve real-world problems. Graduate researcher in CS with math/stat backgrounds.

EDUCATION

01/15/2016 – 08/15/2020 **The School of Computer Science, Carnegie Mellon University**
PhD in Computational Biology — Advisor: Wei Wu

03/01/2011 – 02/26/2013 **Seoul National University**, Korea
MS in Statistics — Advisor: Taesung Park

03/01/2006 – 08/25/2010 **Sungkyunkwan University**, Korea
BS in Mathetmatical Education · BE in Statistics

WORKING EXPERIENCE

Sep 2020 – Present **Applied Scientist**
Amazon.com, Inc., Seattle, WA

Jan 2016 – Aug 2020 **Research Assistant** for Prof. Wei Wu
Computational Biology Department, Carnegie Mellon University, Pittsburgh, PA

May 2018 – Dec 2018 **Research Intern** in Artificial Intelligence and Machine Learning Solution Team
Petuum, Pittsburgh, PA

Aug 2013 – Aug 2015 **Research Assistant** for Prof. Haim Bar
Department of Statistics, University of Connecticut, Storrs, CT

Aug 2011 – Aug 2013 **Research Assistant** in Bioinformatics and Biostatistics
BIBS at Seoul National University, Korea

Aug 2011 – Feb 2013 **Research Assistant** for Prof. Taesung Park
Department of Statistics, Seoul National University, Korea

RESEARCH EXPERIENCE

Interpretable Machine Learning 2018 – 2020
Research Intern advised by Dr. Pengtao Xie — *AI/ML Solution Team at Petuum, Pittsburgh, PA, USA*

- (Worked as an intern at Petuum between May and Dec 2018. Worked as a student at Carnegie Mellon University from 2019.)
- Developed a system-agnostic interpretable ML approach using neural networks using information bottleneck principle.
- Developed an approach to improve an out-of-sample prediction of epitope-TCR binding using an interpretable model.
- Developed a pytorch based software: **VIBI** (github.com/SeojinBang/VIBI) and **TCR** (github.com/SeojinBang/TCR)
- A paper submitted to *[blinded]*

Multimodal/Multiview Machine Learning for Asthma Subtype Identification 2015 – 2020
Research Assistant advised by Dr. Wei Wu — *Carnegie Mellon University, Pittsburgh, PA, USA*

- Developed a multiple kernel k-means clustering approach that is robust against adversarial features.
- Identified variable asthma subtypes by combining multiview clinical data to help clinicians make precision therapy.
- Developed an R-package: **MKKC** (github.com/SeojinBang/MKKC)
- Three papers published in *AJRCCM*, *IEEE BIBM* and *Proteomes*, and a paper submitted to *[blinded]*.

Machine Learning for Natural Language Processing 2018 – 2020

Graduate Researcher *Carnegie Mellon University, Pittsburgh, PA, USA*

- Developed an ontology-based neural network model for patient need detection from an online ovarian cancer discussion forum. A paper submitted to *[blinded]*.
- Analyzed to detect attackable sentences in arguments for successful persuasion using online discussions from the ChangeMyView (CMV) subreddit. A paper submitted to *[blinded]*.

Interpretability and Robustness in Machine Learning 2019

Visiting Graduate Student advised by Dr. Adrian Weller *ML Group at University of Cambridge, UK*

- Investigated the relationship between interpretability and robustness of machine learning and deep learning models.
- Developing a knowledge guided interpretable ML model using the posterior regularization.

Statistical Modeling of Sparse Gene Networks 2013 – 2015

Research Assistant advised by Dr. Haim Bar *University of Connecticut, Storrs, CT, USA*

- Developed a statistical mixture model for better estimating sparse gene network.
- Helped to release an R-package: **edgefinder**
- Submitted a paper to *Statistics in Medicine*.

Statistical Approaches for Biomedical Problems 2012 – 2013

Research Assistant and Scientist advised by Dr. Taesung Park *Seoul National University, Korea*

- Developed a joint feature selection method using the elastic-net regularization to high-dimensional data.
- Deployed a time-dependent survival model to identify subtypes of intraductal papillary mucinous neoplasm.
- Published two papers in *IEEE BIBM* and *Pharmacogenetics and genomics*.

OTHER EXPERIENCE

Professional Experience

2021	Reviewer of *NAACL-HLT, ACM CHIL, ACL/IJCNLP*
2021	Subreviewer of *ISMB*
2020	Reviewer of *NeurIPS, NeurIPS ML4H Workshop, ACM CHIL, ISMB*
2019	Reviewer of *NeurIPS ML4H Workshop*, *IEEE Access*, and *IEEE TNNLS*
2018	Reviewer of *ACM BCB*
2018	Admission Committee at Computational Biology Department, Carnegie Mellon University
2012	Program Committee of two workshops held at Seoul National University and a conference of the Korean Statistical Society
2012 – 2018	Teaching Assistant at Carnegie Mellon University, University of Connecticut and Seoul National University

Leadership Experience

2017 – 2018	Led the communication team in Korean Graduate Student Association to establish and maintain relationship with ∼30 companies seeking to advertise events/job openings.
2016 – 2019	Led the squash team of 35 CMU students. Organized weekly training and trained beginners. Applied and got a financial support from KGSA.

TECHNICAL STRENGTHS

Computer Languages	**Python, R**, C/C++, MATLAB, Bash, HTML
Library & Others	Pytorch, Tensorflow, Caffe, Keras, Git, LaTeX

HONORS AND AWARDS

2018	**The Center for Machine Learning and Health Fellowships in Digital Health**
2013	The Korean Statistical Society Paper Awards (3st Place)

2012 The Korean Statistical Society Poster Awards (1st Place)
2006 – 2010 **National Science and Engineering Undergraduate Scholarship**

PAPERS

[1] **Classifying argumentative relations using logical mechanisms and argumentation schemes**.
TACL (Transactions of the Association for Computational Linguistics), 2021.
Jo, Yohan and **Bang**, **Seojin** and Reed, Chris and Hovy, Eduard.

[2] **Kw-attn: Knowledge infused attention for accurate and interpretable text classification**.
NAACL DeeLIO (The 2nd Workshop on Knowledge Extraction and Integration for Deep Learning Architectures), 2021.
Jang, Hyeju and **Bang**, **Seojin** and Xiao, Wen and Carenini, Giuseppe and Ng, Raymond and Lee, Young ji.

[3] **A mixture model to detect edges in sparse co-expression graphs with an application for comparing breast cancer subtypes**.
*PLOS ONE, **Highlighted Articles***, 2021.
Bar, Haim and **Bang**, **Seojin**.

[4] **Explaining a black-box using deep variational information bottleneck approach**.
AAAI (The Thirty-Fifth AAAI Conference on Artificial Intelligence), 2021.
Bang, **Seojin** and Xie, Pengtao and Lee, Heewook and Wu, Wei and Xing, Eric.

[5] **Robust multiple kernel k-means clustering using min-max optimization**.
preprint arXiv:1803.02458, submitted, 2020.
Bang, **Seojin** and Yu, Yaoliang and Wu, Wei.

[6] **Detecting attackable sentences in arguments**.
EMNLP (Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing), 2020.
Jo, Yohan and **Bang**, **Seojin** and Manzoor, Emaad and Hovy, Eduard and Reed, Chris.

[7] **Identification of epitope-tcr binding using a generative adversarial network model**.
ICML WCB (The 2020 ICML Workshop on Computational Biology), 2020.
Bang, **Seojin** and Lee, Heewook.

[8] **Dropout prediction over weeks in moocs via interpretable multi-layer representation learning**.
AAAI AI4EDU (Workshop on Artificial Intelligence for Education), 2020.
Jeon, Byungsoo* and Park, Namyong* and **Bang**, **Seojin* (co-first)**.

[9] **Phased-lstm based predictive model for ehrs with complex missing patterns.**
preprint, https://www.cs.cmu.edu/~epxing/Class/10708-17/project-reports/project8.pdf, 2019.
Bang, **Seojin** and Yang, Yang and Wang, Yuchuan.

[10] **Multiview cluster analysis identifies variable corticosteroid response phenotypes in severe asthma**.
*AJRCCM (American Journal of Respiratory and Critical Care Medicine), IF 16.49, **Highlighted Articles***, 2019.
Wu*, Wei and **Bang* (co-first)**, **Seojin** and Bleecker, Eugene and Castro, Mario and Denlinger, Loren and Erzurum, Serpil and Fahy, John and Fitzpatrick, Anne and Gaston, Ben and Hastie, Annette and Israel, Elliot and Jarjour, Nizar and Kerr, Sheena and Levy, Bruce Meyers, Deborah and Moore, Wendy and Peters, Michael and Phipatanakul, Wanda and Sorkness, Ronald and Wenzel, Sally.

[11] **Phosphoproteomic analysis of the amygdala response to adolescent glucocorticoid exposure reveals g-protein coupled receptor kinase 2 as a target for reducing motivation for alcohol**.
Proteomes, 6(4), 2018.
Bertholomey, Megan L. and Stone, Kathryn and Lam, TuKiet T. and **Bang**, **Seojin** and Wu, Wei and Nairn, Angus C. and Taylor, Jane R. and Torregrossa, Mary M.

[12] **Naïve bayes ensemble: A new approach to classifying unlabeled multi-class asthma subjects**.
BIBM (2016 IEEE International Conference on Bioinformatics and Biomedicine), 2016.
Bang, **Seojin** and Wu, Wei.

[13] **Joint selection of snps for improving prediction in genome-wide association studies**.
BIBMW (2012 IEEE International Conference on Bioinformatics and Biomedicine Workshops), 2012.
Bang, **Seojin** and Kim, Yong-Gang and Park, Taesung.

[14] **Ethnic variability in the allelic distribution of pharmacogenes between korean and other populations**.
Pharmacogenetics and genomics, 22(12), 2012.
Kim, In-Wha and Im Kim, Kyung and Chang, Hyeu-jin and Yeon, Bora and **Bang**, **Seojin** and Park, Taesung and Kwon, Ji-sun and Kim, Sangsoo and Oh, Jung Mi.