Seojin Bang

860-709-4109 | seojinb@cs.carnegie.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

Aug 2020 (expected) — **The School of Computer Science, Carnegie Mellon University**
PhD candidate in Computational Biology — Advisor: Wei Wu

Feb 2013 — **Seoul National University**, Korea
MS in Statistics — Advisor: Taesung Park

Aug 2010 — **Sungkyunkwan University**, Korea
BS in Mathetmatical Education · BE in Statistics

RESEARCH EXPERIENCE

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. (in progress)

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

- Developed a system-agnostic interpretable ML approach using neural networks using information bottleneck principle.
- Released a pytorch-based software: **VIBI** (github.com/SeojinBang/VIBI) and **TCR** (github.com/SeojinBang/TCR)
- Submitted a paper to *ICLR*

Multimodal/Multiview Machine Learning for Disease Subtype Identification 2015 – 2019
Research Assistant advised by Dr. Wei Wu — *Carnegie Mellon University, Pittsburgh PA*

- 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.
- Got featured in media: CMU News 2019, Pittsburgh Post-Gazette 2019, etc.
- Released an R-package: **MKKC** (github.com/SeojinBang/MKKC)
- Published three papers in *AJRCCM*, *IEEE BIBM* and *Proteomes*, and submitted a paper to *AAAI*.

Machine Learning for Natural Language Processing 2017 – 2019
Graduate Student — *Carnegie Mellon University, Pittsburgh PA*

- Developed an ontology-based neural network model for patient need detection from an online ovarian cancer discussion forum. A paper in progress.

Statistical Modeling of Sparse Gene Networks 2013 – 2015
Research Assistant advised by Dr. Haim Bar — *University of Connecticut, Storrs, CT*

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

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

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] **Explaining a black-box using deep variational information bottleneck approach**.
preprint arXiv:1902.06918, 2020.
Bang, **Seojin** and Xie, Pengtao and Lee, Heewook and Wu, Wei and Xing, Eric.

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

[3] **Detecting attackable sentences in arguments.**
preprint, 2020.
Jo, Yohan and **Bang**, **Seojin**.

[4] **Kw-han: Knowledge infused hierarchical attention network for enhancing interpretability.**
preprint, 2020.
Jang, Hyeju and **Bang**, **Seojin**.

[5] **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.

[6] **A mixture model to detect edges in sparse co-expression graphs**.
Statistics in Medicine, under review, preprint arXiv:1804.01185, 2019.
Bar, Haim and **Bang**, **Seojin**.

[7] **Dropout prediction over weeks in moocs via interpretable multi-layer representation learning**.
AAAI AI4EDU, 2020.
Jeon, Byungsoo and Park, Namyong and **Bang**, **Seojin**.

[8] **Multiview cluster analysis identifies variable corticosteroid response phenotypes in severe asthma**.
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.

[9] **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.

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

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

[12] **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.