

Seojin Bang

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

EDUCATION

08/2015 – 08/2020	The School of Computer Science, Carnegie Mellon University PhD in Computational Biology	Advisor: Wei Wu
03/2011 – 02/2013	Seoul National University, Korea MS in Statistics	Advisor: Taesung Park
03/2006 – 08/2010	Sungkyunkwan University, Korea BS in Mathematical Education · BE in Statistics	

WORKING EXPERIENCE

Sep 2020 – Present	Applied Scientist in ML <i>Amazon.com, Inc., Seattle, WA</i>
Jan 2016 – Aug 2020	Research Assistant <i>Carnegie Mellon University, Pittsburgh, PA</i>
May 2018 – Dec 2018	Research Intern in Artificial Intelligence and Machine Learning Solution Team <i>Petuum, Pittsburgh, PA</i>
Aug 2013 – Aug 2015	Research Assistant <i>Department of Statistics, University of Connecticut, Storrs, CT</i>
Aug 2011 – Feb 2013	Research Assistant <i>Department of Statistics, Seoul National University, Korea</i>

RESEARCH EXPERIENCE

ML for Fraud and Anomaly Detection <i>Applied Scientist</i>	2020 – 2021 <i>Amazon</i>
---	------------------------------

· Blinded

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

- (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 <i>Research Assistant advised by Dr. Wei Wu</i>	2015 – 2020 <i>Carnegie Mellon University, Pittsburgh, PA, USA</i>
---	---

- 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 <i>Graduate Researcher</i>	2018 – 2020 <i>Carnegie Mellon University, Pittsburgh, PA, USA</i>
---	---

- 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 Change-MyView (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*.

PROFESSIONAL SERVICE

2021	Program Committee , <i>NeurIPS XAI4Debugging</i>
2019-2021	Reviewer , <i>NeurIPS, AAAI, ICLR, NAACL-HLT, ACM CHIL, ACL-IJCNLP, ISMB, IEEE Access, IEEE TNNLS, ACM BCB</i>
2018	Admission Committee , <i>Computational Biology Department, Carnegie Mellon University</i>
2011-2012	Program Committee , <i>International Symposium on Statistical Genetics, Microarray Analysis Workshop: Statistical Analysis using R language, The Spring Conference of the Korean Statistical Society</i>

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

PUBLICATION

- [1] **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.
- [2] **Tcr-epitope binding affinity prediction using multi-head self-attention model.**
ICML WCB (The 2021 ICML Workshop on Computational Biology), 2021.
 Cai, Michael and Bang, Seojin and Lee, Heewook.

- [3] **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.
- [4] **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.
- [5] **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**.
- [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, Namyoung* and **Bang, Seojin* (co-first)**.
- [9] **Phased-lstm based predictive model for ehers 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 Bleeker, 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] **Robust multiple kernel k-means clustering using min-max optimization.**
preprint arXiv:1803.02458, submitted, 2020.
Bang, Seojin and Yu, Yaoliang and Wu, Wei.
- [12] **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.
- [13] **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.
- [14] **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.
- [15] **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.