

Seojin Bang

✉ seojinb@alumni.cmu.edu | 🏠 seojinb.com

EDUCATION

2020	The School of Computer Science, Carnegie Mellon University PhD in Computational Biology	Advisor: Wei Wu
2013	Seoul National University, Korea MS in Statistics	Advisor: Taesung Park
2010	Sungkyunkwan University, Korea BS in Mathematical Education · BE in Statistics (double major)	

WORKING EXPERIENCE

Sep 2020 – Present	Applied Scientist <i>Amazon.com, Inc.</i> , Seattle, WA
Jan 2016 – Aug 2020	Research Assistant for Prof. Wei Wu <i>Computational Biology Department, Carnegie Mellon University</i> , Pittsburgh, PA
May 2018 – Dec 2018	Research Intern in Artificial Intelligence and Machine Learning Solution Team <i>Petuum</i> , Pittsburgh, PA
Aug 2013 – Aug 2015	Research Assistant for Prof. Haim Bar <i>Department of Statistics, University of Connecticut</i> , Storrs, CT
Aug 2011 – Aug 2013	Research Assistant in Bioinformatics and Biostatistics <i>BIBS at Seoul National University</i> , Korea
Aug 2011 – Feb 2013	Research Assistant for Prof. Taesung Park <i>Department of Statistics, Seoul National University</i> , Korea

PUBLICATIONS/PREPRINTS

- [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, Namyoung* and **Bang, Seojin* (co-first)**.
- [9] **Phased-lstm based predictive model for ehra 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] **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.**
BIBM (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.

RESEARCH EXPERIENCE

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

- Developed a robust multiple kernel clustering approach.
- Identified asthma subtypes showing differential responses to corticosteroid by combining multiview clinical data.
- Developed an R-package: **MKKC** (github.com/SeojinBang/MKKC)
- Three papers published in *AJRCCM*, *IEEE BIBM* and *Proteomes*, and a paper submitted to [blinded].

Interpretable Machine Learning	2018 – 2020
<i>Research Intern advised by Dr. Pengtao Xie</i>	<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.
- Improved 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]*

Machine Learning for Natural Language Processing

2018 – 2020

Graduate Student

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 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 develop an R-package: **edgefinder**
- A paper under review in *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.
- Two papers published in *IEEE BIBM* and *Pharmacogenetics and genomics*.

PROFESSIONAL SERVICE

2021	Reviewer , <i>NAACL-HLT, ACM CHIL, ACL-IJCNLP</i>
2021	Subreviewer of ISMB
2020	Reviewer , <i>NeurIPS, NeurIPS ML4H Workshop, ACM CHIL, ISMB</i>
2019	Reviewer , <i>NeurIPS ML4H Workshop, IEEE Access, IEEE TNNLS</i>
2018	Reviewer , <i>ACM BCB</i>
2018	Admission Committee , <i>Computational Biology Department, Carnegie Mellon University</i>
2012	Program Committee , <i>International Symposium on Statistical Genetics, Korea</i>
2012	Program Committee , <i>Microarray Analysis Workshop: Statistical Analysis using R language, Korea</i>
2011	Program Committee , <i>The Spring Conference of the Korean Statistical Society, Korea</i>

HONORS AND AWARDS

2018	The Center for Machine Learning and Health Fellowships in Digital Health full tuition and stipend for 12 months and research-related expenses (total \$75,200)	
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 full tuition for 8 semesters an additional \$500 grant for a high GPA an additional \$500 grant for a high GPA	2006 – 2010 2009 2008

SOFTWARE

Python	VIBI: pytorch implementation of VIBI https://github.com/SeojinBang/VIBI
Python	TCR: TCR to epitope binding prediction approach using VIBI https://github.com/SeojinBang/TCR
R-package	MKKC: multiple kernel k -means clustering on a multi-view data https://github.com/SeojinBang/MKKC
LaTeX template	TidyCV: simple and tidy LaTeX template for your curriculum vitae https://github.com/SeojinBang/TidyCV

TECHNICAL STRENGTHS

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

TEACHING EXPERIENCE

Carnegie Mellon University	Teaching Assistant
2018 Quantitative Cell and Molecular Biology Lab	
2017 Computational Methods for Proteogenomics and Metabolomics	
University of Connecticut	Teaching Assistant
2014 Mathematical Statistics	
2014 Introduction to Mathematical Statistics	
2013 Elementary Concepts of Statistics	
2013 Introduction to Statistics I and II	
2013 Statistical Methods	
Seoul National University, Korea	Teaching Assistant
2012 Statistics Laboratory	
2012 Regression and Analysis and Laboratory	
2011 Statistics	
Bongyoung Girls' Middle School, Korea	Student Teacher
2009 Middle School Mathematics	
Sungkyunkwan University, Korea	Teacher
2006 Alternative Elementary/Middle School Mathematics	