ISC, LG Sciencepark, 30, Magokjungang 10-ro
Gangseo-gu, Seoul 07796, Korea
+82 10 9779 7036
sc782@cornell.edu
sc782.github.io

Sungjun Cho

Research Interests

Natural Langauge Processing, Machine Learning, Geometric Deep Learning
Generative Modeling, Computational Linguistics, Cognitive Neuroscience

Education

2018–2020 **M.S. in Computer Science**, Cornell University, Ithaca, NY
- Advisors: David Bindel and David Mimno
- Thesis: Robust and Scalable Spectral Topic Modeling for Large Vocabularies

2011–2017 **B.A. in Computer Science and Mathematics**, Cornell University, Ithaca, NY

Work and Research Experience

Feb 2022 –Present **Research Scientist**, Advanced ML Lab, LG AI Research, Seoul, Korea
- Designed a sparse-attention module that reduces computational cost by data-adaptively choosing its sparsity. Conducted experiments on synthetic token-matching task as well as LRA and GLUE benchmarks.
- Developed a self-supervised molecular pretraining framework with 3D denoising and cross-modal distillation for transferrable molecular representation learning. Conducted experiments on QM9 and OGB benchmarks.
- Applied Riemannian geometry to Transformers to design a non-Euclidean graph Transformer architecture with learnable curvatures. Conducted experiments on graph reconstruction and node classification datasets.
- Participated in other projects on geometric deep learning, continual learning and unlearning, molecular property prediction, image classification, video captioning, music generation, and time-series forecasting.

Sep 2021 –Jan 2022 **Research Intern**, Fundamental Research Lab, LG AI Research, Seoul, Korea
- Proposed a graph pooling module using adaptive number of clusters for molecular graph learning.
- Managed experiments on molecular fluorescence, binding-affinity, and toxicity prediction tasks.

Aug 2020 –Aug 2021 **Graduate Research Assistant**, Computational Science and Engineering, Georgia Tech, Atlanta, GA
- Derived spectral characterization of pathogen load-based 2-mode-SIS model on patient-location networks.
- Developed precautions based on characterization and tested its effect on suppressing spread of MRSA.

Aug 2018 –May 2020 **Graduate Teaching Assistant**, Computer Science, Cornell University, Ithaca, NY
- Led group of $\geq$30 undergraduate TAs as head TA in teaching CS4820: Introduction to Analysis of Algorithms (1 semester) and CS1112/1132: Introduction to Computing using MATLAB (3 semesters).
- Conducted weekly lab/discussion sections and organized grading sessions on assignments and exams.

Aug 2016 –May 2017 **Undergraduate Teaching Assistant**, Computer Science, Cornell University, Ithaca, NY
- Ran weekly office hours and participated in grading sessions for CS2800: Discrete Structures (2 semesters).

Honors and Awards

Nov 2019 **Student Travel Scholarship**, Conference on Empirical Methods in Natural Language Processing

May 2019 **Outstanding Graduate Teaching Assistant Award**, Cornell Computer Science
- For work as graduate teaching assistant for CS4820 and CS1112/1132

May 2017 **Outstanding Undergraduate Teaching Assistant Award**, Cornell Computer Science
- For work as undergraduate teaching assistant for CS2800

Publications

(* denotes equal contribution)

Conference and Journal Papers

Sungmin Cha*, **Sungjun Cho***, Dasol Hwang*, Honglak Lee, Taesup Moon, Moontae Lee. Learning to Unlearn: Instance-wise Unlearning for Pre-trained Classifiers. *AAAI conference on Artificial Intelligence (AAAI).* 2024.

Jiaming Cui*, **Sungjun Cho***, Methun Kamruzzaman, Matthew Bielskas, Anil Vullikanti, B. Aditya Prakash. Using Spectral Characterization to Identify Healthcare-associated Infection (HAI) Patients for Clinical Contact Precaution. *Scientific Reports*. 2023.

Sungjun Cho, Seunghyuk Cho, Sungwoo Park, Hankook Lee, Honglak Lee, Moontae Lee. Mixed-Curvature Transformers for Graph Representation Learning. *Workshop on Topology, Algebra, and Geometry in Machine Learning (TAG-ML at ICML).* 2023.

Sungmin Cha, **Sungjun Cho**, Dasol Hwang, Sunwon Hong, Moontae Lee, Taesup Moon. Rebalancing Batch Normalization for Exemplar-based Class-Incremental Learning. *Conference on Computer Vision and Pattern Recognition (CVPR).* 2023.

Sung Moon Ko, **Sungjun Cho**, Dae-Woong Jeong, Sehui Han, Moontae Lee, Honglak Lee. Grouping-matrix based Graph Pooling with Adaptive Number of Clusters. *AAAI conference on Artificial Intelligence (AAAI).* 2023.

Sungjun Cho, Seonwoo Min, Jinwoo Kim, Moontae Lee, Honglak Lee, Seunghoon Hong. Transformers meet Stochastic Block Models: Attention with Data-Adaptive Sparsity and Cost. *Conference on Neural Information Processing Systems (NeurIPS).* 2022.

Jinwoo Kim, Tien Dat Nguyen, Seonwoo Min, **Sungjun Cho**, Moontae Lee, Honglak Lee, Seunghoon Hong. Pure Transformers are Powerful Graph Learners. *Conference on Neural Information Processing Systems (NeurIPS).* 2022.

Jinwoo Kim, Saeyoon Oh, **Sungjun Cho**, Seunghoon Hong. Equivariant Hypergraph Neural Networks. *European Conference on Computer Vision (ECCV).* 2022.

Moontae Lee, **Sungjun Cho**, Kun Dong, David Mimno, and David Bindel. On-the-fly Rectification for Robust Large-Vocabulary Topic Inference. *International Conference on Machine Learning (ICML).* 2021.

Moontae Lee, **Sungjun Cho**, David Bindel, and David Mimno. Practical Correlated Topic Modeling and Analysis via the Rectified Anchor Word Algorithm. *Conference on Empirical Methods in Natural Language Processing (EMNLP).* 2019.

Preprints

Sungjun Cho, Dae-Woong Jeong, Sung Moon Ko, Jinwoo Kim, Sehui Han, Seunghoon Hong, Honglak Lee, Moontae Lee. 3D Denoisers are Good 2D Teachers: Molecular Pretraining via Denoising and Cross-Modal Distillation. arXiv 2023.

Work In Progress

Seungyeon Rhyu, Kichang Yang, **Sungjun Cho**, Jaehyeon Kim, Kyogu Lee, Moontae Lee. Practical Symbolic Music Generation with Large Language Models using Structural Embeddings.

Byoungjip Kim, Dasol Hwang, **Sungjun Cho**, Honglak Lee, Moontae Lee. Show, Think, and Tell: Learning to Generate Video Captions with Large Language Models.

Minhyuk Seo, Hyunseo Koh, Wonje Jeung, Min Jae Lee, San Kim, Hankook Lee, **Sungjun Cho**, Sungik Choi, Hyunwoo Kim, Jonghyun Choi. Learning Equi-angular Representations for Online Continual Learning.

Jaehoon Lee, Hankook Lee, Sungik Choi, Sungwoo Park, **Sungjun Cho**, Moontae Lee. Periodic and Random Sparsity for Multivariate Long-Term Time-Series Forecasting.

Thesis

Sungjun Cho, Robust and Scalable Spectral Topic Modeling for Large Vocabularies. *M.S. Thesis, Cornell University.* 2020.

Presentations

Jul 2023 Mixed-Curvature Transformers for Graph Representation Learning
- Poster at TAG-ML Workshop at ICML 2023 Conference. Online Virtual.

Mar 2023 Transformers meet Stochastic Block Models: Attention with Data-Adaptive Sparsity and Cost
- Poster at LG Tech Conference. Seoul, Korea.

Nov 2022 Transformers meet Stochastic Block Models: Attention with Data-Adaptive Sparsity and Cost
- Poster at NeurIPS 2022 Conference. New Orleans, USA.

Nov 2022 Transformers meet Stochastic Block Models: Attention with Data-Adaptive Sparsity and Cost
- Poster at 2022 SNU AI Retreat. Seoul, Korea.

Oct 2022 Transformers meet Stochastic Block Models: Attention with Data-Adaptive Sparsity and Cost
- Poster at 1st Yonsei AI Workshop. Seoul, Korea.

Jul 2021 On-the-fly Rectification for Robust Large-Vocabulary Topic Inference
- Poster at ICML 2021 Conference. Online Virtual.

Nov 2019 Practical Correlated Topic Modeling and Analysis via the Rectified Anchor Word Algorithm
- Poster at EMNLP 2019 Conference. Hong Kong, China.

Reviewer Experience

2024 ICLR

2023 ICLR, CVPR, JMLR, ACL, ICCV, NeurIPS