Hyunsoo Cho

Research Interests

Machine Learning

- Anomaly detection, distributional shift, uncertainty estimation, self-supervised learning, unsupervised learning

Natural Language Processing

- Foundation language model, instruction-tuning, imitation learning, prompting, weakly-supervised learning, parameter-efficient transfer learning, black-box language Model

Education

- **Ph.D.** Computer Science and Engineering, Seoul National University **Mar. 2019 — Feb. 2023**
  - **Dissertation:** *A Deep Representation Learning for Unsupervised Anomaly Detection.*
  - Outstanding Ph.D. Dissertation Award
  - **Advisor:** Sang-goo Lee
  - **Committee:** U Kang, Sang-goo Lee, Seung-won Hwang, Sungzoon Cho (SNU), Taeuk Kim (HYU)
- **M.S.** Computer Science and Engineering, Seoul National University **Sep. 2016 — Aug. 2018**
  - **Thesis:** *Visual Question Answering using Simple Natural Language Processing.*
  - **Advisor:** Sang-goo Lee
- **B.S.** Computer Science and Engineering, Chung-ang University **Mar. 2012 — Aug. 2016**

Work Experience

Seoul National University

Postdoctoral Researcher **Mar. 2023 — Current**

Naver Cloud.

Visiting Researcher (Foundation Research Team) **Mar. 2023 — Current**

Theme: HyperCLOVA X (Korean Hyperscale Conversational Language Model)

- Synthetic instruction data generation
- Instruction fine-tuning with curriculum
- Enhancing code capability
- Code capability evaluation

Awards & Honors

Outstanding Ph.D. Dissertation Award

Dept. of Computer Science and Engineering, Seoul National University **Feb. 2023**

Best Poster Award

Fall Retreat, AI Institute of Seoul National University (AIIS) **Nov. 2021**

Publications

Ongoing research & Under review

[1] **Exploring the Correlation between Factual Correctness and Instruction Tuning**
Hyunsoo Cho
Ongoing research

[2] **Instruction Tuning with Human Curriculum**
Hyunsoo Cho*, Bruce W. Lee*,Kang Min Yoo.
** Equal contribution; Preprint, under review.*

International Conferences

[1] **Universal Domain Adaptation for Robust Handling of Distributional Shifts in NLP**
Hyuhng Joon Kim, **Hyunsoo Cho**, Sang-Woo Lee, Junyeob Kim, Choonghyun Park, Sang-goo Lee, Kang Min Yoo, Taeuk Kim.
Findings of the Association for Computational Linguistics: EMNLP 2023 ***(Findings of EMNLP 2023)***.

[2] **Probing Out-of-Distribution Robustness of Language Models with Parameter-Efficient Transfer Learning.**
Hyunsoo Cho, Choonghyun Park, Junyeop Kim, Hyuhng Joon Kim, Kang Min Yoo, Sang-goo Lee.
The 12th Joint Conference on Lexical and Computational Semantics ***(*SEM 2023)***.

[3] **CELDA: Leveraging Black-box Language Model as Enhanced Classifier without Labels.**
Hyunsoo Cho, Youna Kim, Sang-goo Lee.
The 61st Annual Meeting of the Association for Computational Linguistics ***(ACL 2023)***.

[4] **Prompt-Augmented Linear Probing: Scaling Beyond The Limit of Few-shot In-Context Learners.**
Hyunsoo Cho, Hyuhng Joon Kim, Junyeob Kim, Sang-Woo Lee, Sang-goo Lee, Kang Min Yoo, Taeuk Kim.
Thirty-Seventh AAAI Conference on Artificial Intelligence ***(AAAI 2023)***.

[5] **Ground-Truth Labels Matter: A Deeper Look into Input-Label Demonstrations.**
Junyeob Kim, Kang Min Yoo, Hyuhng Joon Kim, **Hyunsoo Cho**, Hwiyeol Jo, Sang-Woo Lee, Sang-goo Lee, Taeuk Kim.
The 2022 Conference on Empirical Methods in Natural Language Processing ***(EMNLP Oral 2022)***.

[6] **Enhancing Out-of-Distribution Detection in Natural Language Understanding via Implicit Layer Ensemble.**
Hyunsoo Cho, Choonghyun Park, Jaewook Kang, Kang Min Yoo, Taeuk Kim, Sang-goo Lee.
Findings of the Association for Computational Linguistics: EMNLP 2022 ***(Findings of EMNLP 2022)***.

[7] **Masked Contrastive Learning for Anomaly Detection.**
Hyunsoo Cho, Jinseok Seol, Sang-goo Lee.
The 30th International Joint Conference on Artificial Intelligence ***(IJCAI 2021)***.

[8] **Automatic Generation of Multiple-Choice Fill-in-the-blank Question Using Document Embedding.**
Junghyuk Park, **Hyunsoo Cho**, and Sang-goo Lee.
*The 19th International Conference on Artificial Intelligence in Education (****AIED 2018****)*.

Workshop

[1] **Self-Generated In-Context Learning: Leveraging Auto-regressive Language Models as a Demonstration Generator.**
Hyuhng Joon Kim, **Hyunsoo Cho**, Junyeob Kim, Taeuk Kim, Kang Min Yoo, Sang-goo Lee
Workshop on Large-scale Pre-trained Language Models ***(LPLM 2022)*** at NAACL 2022.

[2] **Leveraging Class Hierarchy in Fashion Classification.**
Hyunsoo Cho, Chaemin Ahn, Kang Min Yoo, Jinseok Seol, Sang-goo Lee.
The 2nd Workshop on Computer Vision for Fashion, Art and Design 2019 ***(CVFAD 2019)*** *at ICCV 2019*.

[3] **Stochastic Relational Network.**
Kang Min Yoo, **Hyunsoo Cho**, Hanbit Lee, Jeeseung Han, Sang-goo Lee.
The 1st Workshop on Statistical Deep Learning in Computer Vision 2019 ***(SDLCV 2019)*** *at ICCV 2019*.

Domestic

[1] **FastText**를 적용한 한국어 단어 임베딩.
Hyunsoo Cho, Sang-goo Lee.
Korea Software Congress ***(KSC 2017)***.

Patents

- Anomaly Detection System and Method for Setting Threshold Thereof (international - America, China, Korea)

Research Experience

Enhancing Natural Language Understanding via Pre-trained Language Models

Project Manager, with ***Naver Corp.*** **Apr. 2021 — Mar. 2023**

- Prompt optimization.
- Parameter-efficient transfer learning.
- Unsupervised anomaly detection.

An Advanced Study of QA Technologies for Intelligent Assistant System

Project Manager, with ***Hyundai Motors.*** **Mar. 2020 — Feb. 2021**

- Advancing vehicle-embedded intention classifier.
- Detecting anomalous user request from spoken language.

Text and Sentiment Analysis for Fashion Trend Analysis

Project Manager, with ***Korea Creative Content Agency (KOCCA).*** **Jul. 2019 — Jun. 2020**

- Analyzing sentiment in fashion postings through meta information.
- Extracting prominent keywords and entities from fashion postings.

Visiting Talks

- Invited speaker - *Tech talk, **Naver** Enterprise* **Mar. 2022**
- Invited speaker - *Tech talk, **Kakao** Enterprise* **Jan. 2022**
- Invited speaker - *Tech talk, **Hyperconnect** Enterprise* **Jan. 2022**
- Invited speaker - *The 2nd Workshop on Artificial Intelligence for Anomalies and Novelties, **AI4AN 2021** at IJCAI 2021* **Aug. 2021**
- Invited speaker - *2018 토대연구 학술회의 <이보형 民俗樂 Odyssey> **at Korean Music Society*** **May. 2018**

Academic Services

Reviewer:

- The Annual Meeting of the Association for Computational Linguistics (**ACL**): 2022, 2023
- The North American Chapter of the Association for Computational Linguistics (**NAACL**): 2022
- The International Conference On Computational Linguistics (**COLING**) : 2022
- The Conference on Empirical Methods in Natural Language Processing (**EMNLP**): 2022, 2023
- The Association for the Advancement of Artificial Intelligence (**AAAI**): 2023, 2024
- The Joint Conference on Lexical and Computational Semantics (***SEM**): 2023

Teaching Experience

- **Advanced Database** **Fall. 2019**
  - *Department of Computer Science & Engineering, SNU - TA*
- **Introduction to Databases** **Fall. 2017**
  - *Department of Computer Science & Engineering, SNU - TA*
- **IT Entrepreneurship** **Fall. 2017**
  - *Department of Computer Science & Engineering, SNU - TA*
- **Advanced Computer Science Seminar** **Spring. 2017**
  - *Department of Computer Science & Engineering, SNU - TA*