SEYEON (JULIA) LEE

Phone: +8210-3788-7429 Email: seyeonle@usc.edu/cleverlsy@gmail.com https://seyeon-lee.github.io

EDUCATION

Aug. 2019 ~ May. 2021 **University of Southern California** Los Angeles, CA, U.S.A
Viterbi School of Engineering
Master of Science in Computer Science
Advisor: Prof. Xiang Ren

Mar. 2014 ~ Aug. 2018 **Seoul Women's University** Seoul, Korea
Department of Information Media
Bachelor of Engineering in Computer Science & Engineering

Aug. 2016 ~ Jul. 2017 **Illinois Institute of Technology** Chicago, IL, U.S.A
School of Applied Technology
Exchange Student, Information Technology and Business

RESEARCH INTEREST

- **Analysis on Pre-trained Language Models**
- Diagnose the ability in pre-trained language models with newly created probing task
- Devise an efficient method of probing the knowledge learned of pre-trained language models
- Create dataset to improve the performance
- Design pre-training/fine-tuning task
- **Commonsense Reasoning**
- Construct commonsense dataset to compensate NLP models' limitations
- **Building Better Knowledge Injected Language Models**
- Build language model that contains better ability to capture knowledge and robustness by perturbation attacks
- Analyze how much knowledge language model captures and how it can work as knowledge base
- **Multilingual Language Models, Machine Translation**
- Improve model performance in low resource language (e.g. Korean)
- **Natural Language Processing**
- **Machine Learning in general**

RESEARCH EXPERIENCE

Dec. 2019~ Present **INK Lab, University of Southern California** Los Angeles, CA, U.S.A
Research Student Worker, Advised by Professor Xiang Ren

- Working on constructing a challenging commonsense dataset that evaluates the capabilities of Pre-trained Language Models(PTLMs) in making commonsense inferences and the robustness of these inferences to language variations.
- Working on analyzing and evaluating pre-trained multilingual language models and creating multilingual dataset
- Worked on constructing a novel numerical commonsense probing task with a diagnostic dataset consisting of 3K masked word prediction probes.
- Worked on expansive experiments to check whether our novel pre-training objectives improve a pre-trained text-to-text transformer to pack more commonsense knowledge into its parameters without relying on any external resources.

PUBLICATIONS

1. **Common Sense Beyond English: Evaluating and Improving Multilingual Language Models for Common Sense Reasoning**
   in the Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP) 2021
   Bill Yuchen Lin, **Seyeon Lee**, Xiaoyang Qiao and Xiang Ren
   - Evaluate and improve multilingual language models (ML-LMs) to help advance commonsense reasoning (CSR) beyond English
   - Collect the Mickey corpus, consisting of 561k sentences in 11 different languages for analyzing and improving ML-LMs
   - Propose Mickey Probe, a language agnostic probing task for fairly evaluating the common sense of popular ML-LMs across different languages
   - Create two new datasets, X-CSQA and X-CODAH, by translating their English versions to 15 other languages for cross-lingual commonsense reasoning
   - Propose multilingual contrastive pretraining (MCP)

2. ***Birds have four legs?!* NumerSense: Probing Numerical Commonsense Knowledge of Pre-trained Language Models**
   in Empirical Methods in Natural Language Processing(EMNLP) 2020
   Bill Yuchen Lin, **Seyeon Lee**, Rahul Khanna and Xiang Ren
   - Investigate whether and to what extent we can induce numerical commonsense knowledge from PTLMs as well as the robustness of this process
   - Introduce a novel probing task with a diagnostic dataset, NumerSense, containing 13.6k masked-word-prediction probes

3. **Pre-training Text-to-Text Transformers for Concept-centric CommonSense**
   in the International Conference on Learning Representations(ICLR) 2021 (Previous version in SSL@ NeurIPS 2020)
   Wangchunshu Zhou*, Dong-Ho Lee*, Ravi Kiran Selvam, **Seyeon Lee**, Bill Yuchen Lin and Xiang Ren
   - Propose both generative and contrastive objectives for learning common sense from the text to augment PTLMs with concept-centric commonsense knowledge
   - Set generative and contrastive objectives as intermediate self-supervised learning tasks for incrementally pre-training PTLMs
   - Design a joint pre-training framework to unify generative and contrastive objectives to mutually reinforce each other

4. **LEAN-LIFE: A Label-Efficient Annotation Framework Towards Learning from Explanation**
   in Annual Meeting of the Association for Computational Linguistics(ACL) 2020(system demo)
   Dong-Ho Lee*, Rahul Khanna*, Bill Yuchen Lin, Jamin Chen, **Seyeon Lee**, Qinyuan Ye, Elizabeth Boschee, Leonardo Neves and Xiang Ren
   - Introduce LEAN-LIFE, a web-based, Label-Efficient AnnotatioN framework for sequence labeling and classification tasks with an easy-to-use UI

PREPRINT

1. **RICA: Evaluating Robust Inference Capabilities Based on Commonsense Axioms**
   In arxiv
   Pei Zhou, Rahul Khanna, **Seyeon Lee**, Bill Yuchen Lin, Daniel Ho, Jay Pujara, Xiang Ren
   - Propose a new challenge, RICA: Robust Inference capability based on Commonsense Axioms to evaluate the pre-trained language model's ability to make commonsense inferences and how robust the model can be in language variations
   - Develop a systematic and scalable procedure using commonsense knowledge bases and probe PTLMs
   - Automate data generation from first order logic expressions to natural language statements by training sequence-to-sequence model for high-quality statements even from perturbed expressions

WORK EXPERIENCE

May. 2017~ Jul. 2017 **InnerGBUS** Chicago, IL, U.S.A.

Intern, Software Development Division

- Worked with web development team
- Developed websites for a range of divisions across the company using PHP, Javascript, HTML, CSS, AWS
- Contributed to the development of networking software by providing technical support

SKILLS

Python, Pytorch, Transformers, Fairseq, spaCy, scikit-learn
Swift, Java, C++, JavaScript, React, Django, Vue.js

AWARDS AND HONORS

2018	Full Tuition (Academic) Scholarship, Seoul Women's University, Korea
2017	Internship Scholarship, Seoul Women's University, Korea
2016	Exchange Student Scholarship, Seoul Women's University, Korea
2016	SWELL (English Course) Scholarship, Seoul Women's University, Korea

ACTIVITIES

2018	*Volunteer,* teaching math in a national shelter, Seoul, Korea
2018	*Member*, Programming Study Club, Seoul, Korea
2017	*Member*, English Club HaengDang-GMP, Seoul, Korea
2017	*Volunteer*, TED-X, assisting stage management, Chicago, IL, U.S.A.
2014	*Member*, Cheering Squad S.W.U.R.S, Seoul, Korea