

Younghun Lee

Email: younghun@purdue.edu Webpage: <https://youngnns.github.io>

Profile

Computer Science Ph.D. student interested in **natural language processing** and its real world applications, specifically in subjectivity analysis, representation learning, explainable AI, and social network analysis.

Education

| | |
|----------------------|---|
| Aug 2019 to Present | Purdue University , West Lafayette, Indiana Doctor of Philosophy (PhD) in Computer Science <ul style="list-style-type: none">Study Emphasis: Natural Language Processing, Advisor: Professor Dan Goldwasser |
| Sep 2015 to Apr 2017 | University of Michigan , Ann Arbor, Michigan Master of Science in Information <ul style="list-style-type: none">Study Emphases: Information Retrieval and Human-computer InteractionResearch: Interaction retrieval and analysis, advised by Professor Mark W. NewmanAward: Outstanding Graduate Student Instructor of the Year in 2016-2017. |
| Mar 2008 to Feb 2015 | Seoul National University , Seoul, South Korea Bachelor of Science in Electrical and Computer Engineering <ul style="list-style-type: none">Study Emphases: Applied Electrical Engineering and Information VisualizationResearch: Data mining and visualization, advised by Professor Joonhwan LeeAward: Dean of Faculty of Liberal Education Prize (1st Prize) in Fall 2012. |

Selected Publications

| | |
|--------------|---|
| Oct 2018 | Younghun Lee , Seunghyun Yoon and Kyomin Jung, <i>Comparative Studies of Detecting Abusive Language on Twitter</i> , EMNLP 2018 Workshop on Abusive Language Online, Belgium |
| Jul 2020 | Hou Wei Chou, Younghun Lee , Lei Chen, Yandi Xia, Wei-Te Chen, <i>CamemBERT and BiT Feature Extraction for Multimodal Product Classification and Retrieval</i> , SIGIR eCom, China |
| Under Review | Younghun Lee and Dan Goldwasser, <i>Towards Explaining Subjective Ground of Individuals on Social Media</i> , EMNLP 2022 |
| Under Review | Tianqi Wang, Lei Chen, Younghun Lee , Jing Gao, <i>Control False Negative Instances in Contrastive Learning to Improve Long-tailed Item Categorization</i> , AAAI 2022 Main Track |

Professional Experience

| | |
|----------------------|--|
| Jun 2020 to Aug 2020 | Research Science Intern at Rakuten Institute of Technology , Boston, Massachusetts <ul style="list-style-type: none">Conducted research on long-tailed item categorization in natural language processing, improving the state-of-the-art models based on the data augmentation |
| Apr 2018 to Dec 2018 | Research Intern at Seoul National University, Machine Intelligence Lab <ul style="list-style-type: none">Conducted independent research on natural language processing, focusing on abusive language detection on Twitter |

Selected Projects

| | |
|----------------------|--|
| Feb 2017 to Apr 2017 | Revealing the Network and Diffusion Patterns of False Information <ul style="list-style-type: none">Customized <i>Twitter API for Python</i> in order to retrieve more than a million messages containing specific URL information from TwitterModeled friends and follower network graphs, and analyzed them with centrality measures and identified characteristics of false information |
|----------------------|--|

| | |
|----------------------|--|
| Nov 2016 to Apr 2017 | User Interaction Pattern Retrieval and Analysis <ul style="list-style-type: none"> Implemented an Android web emulator from a Java application that captures user interaction patterns in mobile applications and stores log data in Firebase database Recruited 100 users from Amazon Mechanical Turk and analyzed log files in order to determine the type of tasks |
| Nov 2016 to Dec 2016 | Investigating the Effect of Presidential Election Results on Hate Crimes <ul style="list-style-type: none"> Developed a Python program with NLTK in order to retrieve Twitter messages on hate crimes and analyze the sentiment of each message Investigated the statistical significance of a null hypothesis with t-statistics |
| Mar 2016 to Apr 2016 | Developing a UFO Search Engine (Summary Paper) <ul style="list-style-type: none"> Improved an existing UFO search engine by evaluating different information retrieval methodologies such as inverted index, tf-idf, PageRank, and query expansion |
| May 2014 to Jun 2014 | Sawtooth Bubble Chart—Visualizing Human Emotion (Summary Paper) <ul style="list-style-type: none"> Retrieved and categorized Twitter data by time, age, and geolocation Implemented a model that visualizes arousal and valence level of human emotion of each tweet using Affective Norms for English Words dictionary |
| Mar 2014 to May 2014 | Korean Text Analysis—Relevancy Visualization of Morphemes <ul style="list-style-type: none"> Developed a vector space model that parses a Korean document, calculates the intensity of relevancy between meaningful words by cosine similarities Improved the model by experimenting with weights on different n-grams |
| Dec 2013 to Jan 2014 | Visualizing Articles with Different Viewpoints <ul style="list-style-type: none"> Built a web application that parses two Korean articles of the same subject into morphemes, classifying their similarities and differences, and visualizing them with word clouds |

Invited Talks

| | |
|----------|---|
| Dec 2018 | “Comparative Studies on Detecting Abusive Language on Twitter” at NAVER [link] |
|----------|---|

Reference

| | |
|-----------------------|---|
| Dan Goldwasser | Associate Professor <ul style="list-style-type: none"> Purdue University, Department of Computer Science 2142A Lawson Bldg, 305 N University St, Lafayette, Indiana 47907 Email: dgoldwas@purdue.edu |
| Mark W. Newman | Associate Professor <ul style="list-style-type: none"> University of Michigan, School of Information 4380 North Quad, 105 S State St, Ann Arbor, Michigan 48105 Office: +1 (734) 764-0020 Email: mwnewman@umich.edu |
| Kentaro Toyama | W.K. Kellogg Associate Professor of Community Information <ul style="list-style-type: none"> University of Michigan, School of Information 3444 North Quad, 105 S State St, Ann Arbor, Michigan 48105 Office: +1 (734) 763-8427 Email: toyama@umich.edu |