

Taehee Jung

Pittsburgh, PA 15232

☎ 412-403-8832 | ✉ taj41@pitt.edu | 🌐 <http://theejung.github.io>

Education

University of Pittsburgh

Pittsburgh, PA, USA

PH.D., DEPARTMENT OF STATISTICS

Aug 2018 - Present

- Advisor: Lucas Mentch (UPitt Stats), closely working with Eduard Hovy (CMU LTI)

University of California, Berkeley

Berkeley, CA, USA

MASTER OF ARTS, DEPARTMENT OF STATISTICS

Aug 2016 - May 2017

Korea Advanced Institute of Science and Technology (KAIST)

Daejeon, Korea

BACHELOR OF ARTS, DEPARTMENT OF MANAGEMENT SCIENCE

2007 - 2011

- Academic excellence scholarship (2007 - 2011)

Professional Experience

3M M*Modal

Pittsburgh, PA

RESEARCH NLU INTERN

Summer 2019

- Collaborators: Thomas Schaaf
- Reinforcing posterior calibration optimization to sequence predictions tasks such as Part-of-speech (POS) tagging and Named Entity Recognition (NER). This work is in preparation to submit at [ACL 2020].
- I interviewed with 3M Career website with respect to the internship experience. Please click this [link](#).

University of Pittsburgh

Pittsburgh, PA

GRADUATE RESEARCH ASSISTANT

Dec 2018 - Present

(1) ANALYZING CORPUS AND SYSTEM BIASES IN CURRENT SUMMARIZATION

Spring 2019

- Collaborators: Eduard Hovy (CMU LTI), Lucas Mentch (UPitt Stats), Dongyeop Kang (CMU LTI)
- I analyzed how current corpora or summarization systems are biased toward certain sub-aspects (i.e., position of the sentences, importance, and meaning diversity). This work will appear in [EMNLP 2019].

(2) GEOFITTING: CONSTRAINING RELATIONAL SEMANTICS INTO WORD EMBEDDINGS

Fall 2018

- Collaborators: Eduard Hovy (CMU LTI), Dongyeop Kang (CMU LTI)
- I analyzed geometric tendencies (e.g., slope, distance) on different set of inter-world relations, and observed the non-linearity of relations on existing word vectors (i.e. Word2Vec, Glove). Then, I developed an optimization technique called *geo-fitting* that minimizes the geometric differences but keeps the properties from the original word vectors, and showed that the geometry constrained vectors help many downstream NLP tasks (e.g., semantic similarity, textual entailment). This work is submitted to [AAAI 2020].

University of California, Berkeley

Berkeley, CA

(1) EMPIRICAL EXAMINATION ON GENERATIVE ADVERSARIAL MODELS, ML CLASS PROJECT [REPORT]

Feb-May 2017

- Studied/implemented different types of adversarial generative models: vanilla, deep convolutional, Wasserstein, and Cycle GAN.
- Compared them on different tasks: image generation, interpolation, projection, arithmetic operations, translation, and completion.
- Won the best credit among Berkeley's Machine Learning course projects.

(2) S&P 500 STOCK PORTFOLIO WITH SENTIMENT ANALYSIS, CAPSTONE PROJECT [REPORT] [PPT]

Jan - May 2017

- Analyzed sentiment of tweets predicting volatility and developed autoregressive predictive models
- Constructed a stock portfolio using Markowitz objective function with L1 penalty

(3) CREATING R PACKAGE FOR MFA, CLASS PROJECT OF STATISTICAL COMPUTING [DEMO]

Nov - Dec 2016

- Developed R package for Multiple Factor Analysis to perform MFA, bootstrapping and visualized the results on Shiny app

Korea Credit Bureau (KCB)

Seoul, South Korea

DATA SCIENTIST

Dec 2010 - May 2016

(1) BIG, PUBLIC DATA ANALYSIS, FUNDED BY MINISTRY OF GOVERNMENT ADMINISTRATION & JEONJU CITY

Sep - Dec 2015

- Analyzed and visualized boarding histories, local floating population, petition documents, and home to office O/D information
- Constructed new strategies for bus rearrangement and adopted by Jeonju city on Feb, 2017
- Practically chosen to Jeonju city's bus rearrangement plan [article] [report] (written in Korean)

(2) FRAUD DETECTION SYSTEM DEVELOPMENT, FUNDED BY WOORI BANK

Dec 2014 - Aug 2015

- Segmented customers' transaction histories and developed predictive ensemble models on regression, neural net, and decision tree
- Compared with previous FDS, overall detection rate increases more than 20% [article] (written in Korean)

(3) MIGRANT ANALYSIS FOR MAGOK NEW TOWN, FUNDED BY LG ECONOMIC RESEARCH INSTITUTE

Jun - Sep 2014

- Developed demographic database of migrants and classified them using decision tree for potential cultural demand to the new town

(4) LOCATION BASED DECISION SUPPORT SYSTEM, FUNDED BY KOREA CREDIT BUREAU & R114

May - Aug 2012

- Developed an integrated database (e.g., location, real estate, residential's demographic and financial data) called K-atlas

Publication

WORK IN SUBMISSION OR PROGRESS

Don't Treat Every Relation the Same! Constraining Word Vectors to Geometric Properties of Inter-Word Relations, Taehee Jung*, Dongyeop Kang*, Eduard Hovy (*equal contribution), Submitted to AAAI 2020

Posterior Calibrated Training on Sequence Prediction Tasks, Taehee Jung*, Lucas Mentch, Thomas Schaaf, In preparation to submit at ACL 2020

PUBLISHED PAPERS

EMNLP'19 **Earlier Isn't Always Better: Sub-aspect Analysis on Corpus and System Biases in Summarization**, Taehee Jung*, Dongyeop Kang*, Lucas Mentch, Eduard Hovy (*equal contribution), Conference on Empirical Methods in Natural Language Processing (EMNLP), 2019

Academic Experience

Invited for women PhD candidates and researchers

Menlo Park, CA

FACEBOOK'S WOMEN IN RESEARCH LEAN IN.

Sep 11 - 13, 2019

Teaching Assistant at University of Pittsburgh

Pittsburgh, PA

STATISTICS & PROBABILITY FOR BUSINESS MANAGEMENT

Spring 2019

APPLIED STATISTICAL METHODS

Fall 2018

Service

2019 **External Reviewer**, Annual Meeting of the Association for Computational Linguistics (ACL)

2019 - Present **Lead Organizer**, Statistics and Machine Learning Group, University of Pittsburgh

Pittsburgh, PA

Skills

Programming Python, R, SAS, SQL, Unix/Linux, Tableau, QGIS

Language English (Fluent), Korean (Native), Spanish (Beginner)