Jeffrey (Young-Min) Cho

Phone: (857) 576-7111
Email: jch0@seas.upenn.edu
cym940514@gmail.com
Website: https://jeffreych0.github.io/

RESEARCH INTERESTS

Natural Language Processing, Artificial Intelligence, Machine Learning, Conversational Agents, Large Language Models, Natural Language Reasoning, Computational Social Listening.

EDUCATION

Ph.D in Computer and Information Science
University of Pennsylvania, Philadelphia, PA
2023 - Present

MSE in Data Science
University of Pennsylvania, Philadelphia, PA
Graduation - 2022

BS in Economics, Minor in Applied Statistics
Yonsei University, Seoul, Korea
Graduation - 2020

PUBLICATIONS

[6] C. A. Stamatis, T. Liu, J. Meyerhoff, Y. Meng, **Y. Cho**, C. J. Karr, B. L. Curtis, L. H. Ungar and D. C. Mohr. *Specific associations of passively sensed smartphone data with future symptoms of avoidance, fear, and physiological distress in social anxiety*. To appear in Internet Interventions, 2023.
[5] **Y. Cho**, S. Rai, L. Ungar, J. Sedoc, S. C. Guntuku. *An Integrative Survey on Mental Health Conversational Agents to Bridge Computer Science and Medical Perspectives*. To appear in EMNLP 2023.
[4] X. D. Hou, S. C. Guntuku, **Y. Cho**, G. Sherman, T. Zhang, M. Li, L. Ungar, L. Tay. *A Cross-cultural Examination of Temporal Orientation Through Everyday Language on Social Media*. To appear in PLOS ONE, 2023.
[3] J. Sohn, S. Jeong, **Y. Cho**, T. Park. *Functional Clustering Methods for Binary Longitudinal Data with Temporal Heterogeneity*. In Computational Statistics & Data Analysis, 2023
[2] **Y. Cho**, L. Zhang and C. Callison-Burch. *Unsupervised Entity Linking with Guided Summarization and Multiple Choice Selection*. In EMNLP 2022.
[1] S. Lim, S. Kwan, **Y. Cho**, T. Park, B. Colvard, A. d'Audiffret, V. Kashyap and J. S. Cho. *Discrepant Effects of Case Volume on Mortality After Elective and Ruptured Abdominal Aortic Aneurysm Repair*. In Journal of Vascular Surgery, 74(3), p.e65. 2021

AWARDS

Solomon M. Swaab Fellow for the 1st Year Ph.D. Student (2023)

Research Experience

8/2023 - present **English Tutoring Chatbot**[5]

Building a chatbot that allows foreign students to practice their conversational English

Advised by Dr. Lyle Ungar

- Use GPT to generate English text at the appropriate level for each student in conversations that incorporate desired topics, vocabulary, and grammatical constructions.
- Explanations can be generated as needed, taking into account cultural background and local languages such as Hindi and Korean.

8/2021 - 6/2022 **Retrieval-Augmented Generation for Reasoning and Entity Linking**[2]

"This is a good band." "Which band? Musical band, ribbon band or wedding band?"

Advised by Dr. Chris Callison-Burch

- Entities are hard to be linked without a definition, especially with an ambiguous context.
- We suggest a dataset designed for complex entity tracking in procedures.
- Jointly using local and global context can boost a model's understanding of input text.
- Entity Linking can be solved by transforming it to a multiple-choice problem, where choices are made by entity relationships in a DB.

5/2021 - 5/2022 **Sense-Sensitive Lexicon Induction**

Generating word clusters handles polysemy and domain difference

Advised by Dr. Lyle Ungar

- Human generated lexica often suffer from a trade off between coverage and polysemy. Some words might be more frequently used with other meaning on the target corpus.
- Token embeddings of polysemous words can be grouped by different senses using iterative clustering and in-cluster monosemy merging.

1/2021 - 6/2021 **Domain Space Alignment for Text Classification**[a]

Domains are latent spaces, and matching these spaces can help domain adaptation

Advised by Dr. Dan Roth

- A simple but effective domain transfer algorithm on embedding for Multi-Domain Text Classification.
- Using Household Transformation to map target sentence embedding to source space.
- The matched embedding gave 1% higher accuracy than original embedding, and achieved 2.5% and 1.6% increase in accuracy for one-shot and two shot cases.

1/2021 - present **Psychological Language Analysis**[4, 6, b, c, d]

Social media can highlight the cultural and psychological variations among different groups

Advised by Dr. Lyle Ungar, Dr. Sharath Chandra Guntuku

- Analysis of language shows that psychological richness is a distinct aspect of a fulfilling life, as evidenced by examining Facebook messages for its discriminant and face validity.
- Twitter and Weibo posts reveals the differences in valence, arousal and temporal orientation between the US and China.

WORK EXPERIENCE

8/2022 - 7/2023 **Chief Data Scientist**[6, d]
World Well-Being Projects, University of Pennsylvania, Philadelphia, PA

- Led data acquisition and analysis from social media and phone sensor sources, utilizing machine learning and NLP to derive insights on well-being and health outcomes.
- Oversaw development projects and mentored programmers and students, driving data refinement and ensuring algorithm and coding quality.

1/2021 - 7/2022 **Research Assistant**[4, c]
World Well-Being Projects, University of Pennsylvania, Philadelphia, PA

- Led analysis and research analyzing social media data to discover cultural and psychological variations between people from different groups.
- Studied methods to enhance human generated lexica using language models.

2/2019 - 1/2021 **Data Scientist**
AmorePacific, Seoul, Korea

- Based on purchase history, predicted Customer Lifetime Value (CLV) using Louvain method and designed a promotion plan for 1000+ Innisfree offline stores in Korea.
- Based on product descriptions and customer reviews, built a model for personalized cosmetics preference.

3/2018 - 3/2019 **Data Analyst Intern**
Smartstudy ("Baby Shark"), Seoul, Korea

- Designed a sales boosting model in China using economic monopoly model. Activated three sales plans by launching product bundle resulting in maximal 86.3% sales boost in single launch.

6/2015 - 4/2016 **Business Developer Intern - Analysis Manager of National Competitor Tracking Team**
Uber Technologies, Inc., Qingdao & Beijing, China

- Led a national team to analyze competitor's market share and translated data into valuable information; drafted detailed reports to support Headquarter's strategy.

PREPRINTS

[d] C. A. Stamatis, J. Meyerhoff, Y. Meng, Z. C. C.Lin, **Y. Cho**, T. Liu, C. J. Karr, T. Liu, B. L. Curtis, L. H. Ungar and D. C. Mohr. *Differential temporal utility of passively sensed smartphone features for depression and anxiety symptom prediction: A longitudinal cohort study*. Conditional Acceptance with npj Mental Health Research.

[c] C. Bonner, **Y. Cho**, F. Zhang, L. Tay, L. Ungar and S. C. Guntuku,. *The Assessment of Psychological Richness, Meaning, and Happiness with Social Media Text Data.* Under Preparation.

[b] **Y. Cho**, S. Thapa, D. Pang, G. Sherman, L. Ungar, L. Tay and S. C. Guntuku. *Cultural differences in Valence Arousal Experience across the United States and China.* Under Preparation.

[a] **Y. Cho**, Y. Zhang, M. Scharf and D. Roth. *Domain Adaptation with Household Transformation for Text Classification.* Under Preparation.

LANGUAGES

English: proficient **Chinese:** native **Korean:** native