

Yi Han

Want to save some time? Personal website: <https://jadehanyi.com/>

(781) 922-6520 | han.yil@northeastern.edu

<https://www.linkedin.com/in/yi-han-2900a1137/> | <https://github.com/hanyidaxia?tab=repositories>

EDUCATION

Northeastern University

Boston, MA

Ph.D. in Industrial Engineering

Expected Dec. 2023

Relevant courses: Deep Learning, Natural Language Processing, Machine Learning, Algorithms, Prob & Stats

Northeastern University

Boston, MA

M.S. in Data Analytics Engineering; GPA: 3.86

Dec. 2018

Relevant courses: Database Design, Data Mining, Data Visualization, Probabilistic Operation Research

SKILLS & AWARDS

Programming:	Python, R, MATLAB, C++, JAVA, Tableau, HTML, CSS, JavaScript
Database & Cloud:	SQL, AWS, SAS, Access, SPSS, MiniTab, Keylines, IOR, LINGO
Graphic Modeling:	Gephi, Anychart, AUTO-CAD, POWEBI, FLUENT, GAMIT
Certificate:	Mooc certificates of ML (2020); DL on Coursera (2020)
Awards:	IDETC-CIE 2022 Hackathon, Problem 3, 3 rd place
Patent:	Multi-Modal Data-Driven Design Concept Evaluator

WORK EXPERIENCE

Lira

NC

AI Tech Lead

Dec. 2022 - Jun. 2023

Task Introduction in short words: **Reading lip moves and translate to text from silent video.**

- Data Science Part Lead the lip reading project
 - Design several models for the lip-reading task, including Transformer-backbone-based, ResNet-backbone-based, and several pure-video based
 - Build the model, in the latest work, and subtraction based transformer-backbone model achieved the state of the art performance in the lip reading task
 - Deploy the model to the app platform (in process)
- Data Engineering Part Lead the database building and data annotation:
 - Design the web for collecting lip reading dataset
 - Build and manage the **Azure** storage for the collected data

Merck

Cambridge, MA

Data Scientist Intern

Jan. 2022 - Jul. 2022

Task Introduction in short words: **Identifying protein related findings in documents to specific disease.**

- Built target liability assessment text analyzing model (includes a **search algorithm** and a **text classifier**) based on the paper of target (compound) searching results from PubMed
 - Capabilities of the model:
 - Extracting all the sentences related to the compound and customizable disease or symptoms
- Built front-to-end prototype of the **deep-learning** based compound analyzing model (**DCM**) for Merck historical compound pdf-format-based reports
 - Capabilities of the model:
 - Parse all the pdf and word files, extracting different sections from those files including the abstract, conclusion, and result
 - Perform the **NER** (name entity recognition) task from the extracted sections

BaiRong Financial Information Service Company

Beijing, China

Machine Learning Intern

Jun. 2018 - Sep. 2018

Task Introduction in short words: **Regression model with XGBoost and LightGBM**

- Created model for risk control through **Logistic regression** and **Stepwise regression** via R and Python; “bad

customer” ratio decreased by 5.67%, and payment received ratio increased by 20.3%

- Stacked multiple **XGboost** into single model, compared with **LightGBM**; increased AUC from 0.68 to 0.76

RESEARCH

Ph.D. Research

Main Research: Latent Needs Elicitation

Mar. 2020 – Present

Objective: Eliciting latent customer needs from customer reviews

Latent Needs Elicitation Through Aspect Category Sentiment Opinion Extraction

Mar. 2023 – Present

- Built a new position encoding algorithm, a new loss function incorporate cross entropy and KL divergence for multi-label, weight-wised task
- Created a **unified deep-learning sequence-to-sequence model** with the new position encoding and Loss which **BEAT** pure T5 on our dataset

May. 2022 – Feb. 2023

- Built an annotated dataset, including labels A (aspect), C (category), O (opinion), S (sentiment) and I (opinion implicit indicator)
- Created a **unified deep-learning sequence-to-sequence model** to extract all labels parallelly based on **T5**
- Conducted clustering analysis for opinion and aspect, identified most contradictory opinion with same aspect

Latent Needs Elicitation Through Aspect-Sentiment Guided Opinion Summarization

Sep. 2021 – Aug. 2022

- Built a sequence-to-sequence MAS-T5 model for the aspect and sentiment-oriented summarization of reviews with Pytorch
- Designed a hierarchical **max-pooling model** MAS, which can predict the sentiment and aspect label in word, sentence, and review level
- Utilized the output from MAS to assemble a **synthetic** supervised summarization data, which can be used for abstract summarization task
- Fine-tuned the **T5** model with the synthetic data

Latent Needs Elicitation Through Aspect, Opinion, Sentiment Extraction

Jun. 2020 - Apr. 2021

- Built a **BERT-NER** model for elicitation of customer needs based on online reviews with Pytorch
- Designed a highly weighted loss function to resolve the extremely unbalanced dataset
- Labeled and assembled the output from the BERT-NER as user needs simulation
- Utilized **BLUE** score to evaluate the results of the needs
- Developed a web crawler to obtain source data
- Designed a double layer of **CNN** on top of BERT as a post-training parallel comparison

Latent Needs Elicitation Through Sentiment and opinion extraction

Mar. 2019 - Aug. 2019

- Built a **data crawler** to organize the original dataset
- Built a **product attribute lexicon** for further analysis (sneaker lexicon)
- Designed two types of algorithms for the **attribute level sentiment analysis**
- Conducted clustering analysis of the customer expression based on the sentiment analysis results

Other Research:

Algorithm Course Design in the College of Engineering for Northeastern University

Jan. 2021 - May. 2021

- Designed course content based on two textbooks *Algorithm Design* and *Algorithms*
- Built the course example code and course quizzes
- Developed exercises based on textbooks
- Drew graph demonstration for classic algorithms like recursive

Unsupervised Attribute Clustering Analysis Based on Customer Reviews

Jan. 2021 - Aug. 2021

- Filtered and clustered critical product attributes with product description via Pytorch
- Conducted clustering analysis based on filtered attributes instead of product

PUBLICATIONS

• Journal Paper

- Han, Y. Moghaddam, M.(2023) Domain knowledge as attention fixer in Large Language Models *Journal of Engineering Design (Under Review)*
- Shi, J., & Yi, H.(2023). Aspect Guided Abstractive Summarization for Safety Concern Information

Extraction. *JCISE (Under review)*

- Han, Y., Nanda, G., & Moghaddam, M. (2022). Attribute-Sentiment-Guided Summarization of User Opinions from Online Reviews. *J. Mech. Des.*, 1–41. doi: 10.1115/1.4055736
- Han, Y., & Moghaddam, M. (2021). Eliciting Attribute-Level User Needs From Online Reviews With Deep Language Models and Information Extraction. *J. Mech. Des.*, 143(6). doi: 10.1115/1.4048819
- Han, Y., & Moghaddam, M. (2021). Analysis of sentiment expressions for user-centered design. *Expert Syst. Appl.*, 171, 114604. doi: 10.1016/j.eswa.2021.114604
- The feasibility study of fire emergency evacuation in the integrated transport system-Beijing south railway station, *China Chemical Trade* (ISSN:1674-5167)

- **Conference Paper**

- A Design Knowledge Guided Position Encoding Methodology for Implicit Need Identification From User Reviews, **IDETC/CIE, 2023**
- A Priori: Design Knowledge in AI, **DesForm, 2023**
- Extracting latent needs from online reviews through deep learning based language model, **ICED, 2023**
- Aspect-Sentiment-Guided Opinion Summarization for User Need Elicitation From Online Reviews, **IDETC/CIE 2022**