Soyoung Yoon

University Address) 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea
+82(42)-350-2114 (school) | soyoungyoon@kaist.ac.kr
github: soyoung97

RESEARCH INTEREST

Natural Language Processing, Machine Learning, Software Engineering

RESEARCH EXPERIENCES

Individual study at KAIST U&I lab (https://uilab.kaist.ac.kr) Sep. 2019 ~

Advisor: Alice Oh (https://aliceoh9.github.io/)
Participate on machine learning reading group

News trend analysis Fall 2019

Identified top ten most significant issues for each year and rank them from news articles over the period of three years. (2015 – 2017) Implemented by pandas, genism, and nltk.
Conducted topic modeling by LDA. Promoted topic quality by neuroNER (Implemented methods suggested by https://www.aclweb.org/anthology/P18-2040/)
Report: https://github.com/soyoung97/Topic_modeling-Issue_Tracking/blob/master/report/report.pdf

Replicate & modify LM (https://github.com/soyoung97/awd-lstm-gru) Fall 2019

Replicated *Regularizing and Optimizing LSTM Language Models* (https://arxiv.org/abs/1708.02182)
Modification by replacing LSTM model to BiLSTM and GRU, each gaining improvement on validation loss and training time.
Poster: https://github.com/soyoung97/awd-lstm-gru/blob/master/poster.pdf

URP(Undergraduate Research Project) Jan. 2020 ~

Research on Grammatical Autocorrection for Korean via fine-tuning pre-trained Language Models
Proposal for the research(In Korean): https://soyoung97.github.io/urp.pdf

Main contribution:

- Trained models (bart, transformer, lstm) using fairseq library
- Made naïve-transformer model using pytorch
- Replicated copy-attention model for Korean (https://www.aclweb.org/anthology/N19-1014.pdf)
- Preprocessed and tokenized Korean data using numpy, pandas, hgtk, and sentencepiece

Advisor: Sungjoon Park(https://sungjoonpark.github.io), Alice oh

Research Intern, Naver Clova AI, Conversation Team July. 2020 ~ Oct.2020

(In progress) Supervised under Gyuwan Kim (https://gyuwankim.github.io/)

EDUCATION

University

Korea Advanced Institute of Science and Technology (**KAIST**), Daejeon, South Korea

Bachelor of Computer Science Feb. 2016 ~ Feb. 2021 (expected)

(Advanced major on Computer Science, Completed courses focused on AI)

Major course

Discrete Mathematics, Data Structure, Programming Principles, System Programming, Linear Algebra, Computer Organization, Algorithms, Operating Systems, Computer Networks, Computer Vision, Programming Languages, Text mining, Machine Learning, Natural Language Processing, Introduction to AI, Data Analytics using R

GPA

Total: 3.77/4.3, (Major only: 3.92/4.3)
GPA of each semester: 3.37(Spring 2016) -> 3.17 -> 3.6 -> 3.78 -> 4.13 -> 3.77 ->4.06 -> 4.3

English Proficiency Tests

TOEFL, 108 Nov. 2018

TOEIC, 975 June. 2020

MAJOR PROJECTS

Pintos (KAIST Course Project) Spring 2018

I went through Pintos projects for the Operating System course in Spring 2018.

MadCamp (https://www.madcamp.io) Summer 2018

I conducted 5 projects per week, mainly focusing on mobile app development, web programming based on Flask, Gan-based machine learning projects using Tensorflow.

I used GAN to transform black&white images to RGB images.

Model code implementation:
(https://github.com/soyoung97/madcamp3/blob/master/colorization_test.py)

Implement javascript library (https://github.com/soyoung97/check) July. 2019

Check is a javascript opensource utility library for debugging, logging, assert, and data checking.

Software Engineer Intern, AITRICS (https://www.aitrics.com) Jan. 2019 ~ Aug. 2019

I worked as a front-end and back-end server engineer, counducting hundreds of data per minute, using Django-Rest Framework and Vue.js.

Major contributions:

-Profiled the speed of services by using Django-silk and Django-debug-toolbar. Optimized SQL queries for big amounts of input data using Django ORM.

-Implemented functions (Giving alarms at appropriate time, Calculating individual patient's medical score using celery and celery-beat) needed for the VitalCare product dashboard UI

-Made current products service well with newly made machine learning models

-Applied Recurrent Neural Network models to the inference server and computed the cosine similarity between real-word and trained data distributions

-Implemented test functions using pytest and python hypothesis

-Implemented dashboard UI by Vue.js (front-end)

-Mainly implemented patient's vital sign graph by using echarts (front-end)

-In charge of the front-end and back-end part of the VitalCare product

Site: https://www.aitrics.com/

KENS (KAIST Course Project) Spring 2020

KENS stands for KAIST Educational Network System. I built prototypes of Ethernet/ARP/IP/TCP, implement congestion control, and contributed to the KENS official repository. (https://github.com/ANLAB-KAIST/KENSv3)

ACCOMPLISHMENTS

Won the third Prize at 2020 winter/spring URP workshop Aug. 2019

Research title: Grammatical Autocorrection for Korean via Fine-tuning pre-trained Language Models

Won ~$700

Conference presentation talk at PyCon Korea 2019 Aug. 2019

Title: Django query optimization for medical AI data processing

Slide: https://www.slideshare.net/SoyoungYoon11/pycon-presentation-final

Video: https://www.youtube.com/watch?v=HpMYWk566OA

Received National Graduate Science & Technology Scholarship 2018

Won ~$500 for a semester (Merit-based scholarship)

Finalist @ Power of xx ctf http://www.powerofcommunity.net/ 2018

Participated as one of the members of Power of xx team.

Conducting Mock Hacking Outsourcing for Hunesion Products 2018

Found two important vulnerabilities in their secure software and provided them with solutions.

Finalist @ Power of xx ctf http://www.powerofcommunity.net/ 2017

Joined KAIST GoN team http://gon.kaist.ac.kr 2017 ~ Present

GoN is a KAIST hacking & security club that study hacking and participate in CTF.