360 Huntington Avenue, Boston, MA ⋄ https://snoopy666.github.io
8579305269 ⋄ wang.mengyua@husky.neu.edu ⋄ Available for 2020 Spring Coop / 2020 Summer Intern

EDUCATION

Northeastern University, Boston, MA Sept 2018 - Dec 2020
M.S. in Computer Science GPA: 3.9/4.0
Selected Courses: Program Design Paradigm, Algorithms, Machine Learning, Special Topics in Deep Learning

Beijing Information Science & Technology University Sept 2014 - June 2018
B.Eng. in Computer Science and Technology

TECHNICAL SKILLS

Computer Languages	Java, Python, C/C++, C#, Javascript, HTML5/CSS, SQL, Bash
Tools & Skills	Git, Docker, MySql, Android, IntelliJ, Matlab, QT, ,Visual Studio, Latex, Vim
Platform & Frameworks	Hadoop, Spark, Hive, AWS, Tensorflow, Pytorch, Scikit-learn, JUnit, React

EXPERIENCE

Meituan-Dianping July -Sept 2019
Software Engineering Intern (Machine Learning) @ Search and NLP Team *Shanghai, China*

Project: Optimize quality and attraction of summarization of users' reviews for each shop on the search page.

- Product Cognitive Iteration: Wrote Hive Sql scripts and UDF functions implemented with Java to extract and analyze billion-level UGC from the database. Used Scala for preprocessing and feed processed data to new tables in database for further analysis. Analyzed results and proposed methodology and metrics for service improvement.
- Explored several task formulation possibilities, e.g. summarization generation under Seq2Seq model with attention mechanism, Pointer-Generator Networks model. Eventually formulated the task as a sequence tagging problem.
- Implemented a hierarchical LSTM models to extract users' reviews as summaries. Trained, tuned model and implemented batch prediction for full and incremental data on Hadoop YARN distributed clusters.
- To obtain higher-level text features, conducted several experiments by modifying and fined-tuning pretrained Bert model and Transformer model to summarize, which achived higher ROUGH score and human evaluation results.
- The model increased Unique Visitor CTR by 6 bp in A/B test and is deployed online for full service.

SELECTED PROJECTS

Coherent Lyrics Generation Conditioned on Melody and Artist Jan - April 2019
Project of Special Topics in Deep Learning Course

- Designed and implemented a novel hierarchical model using LSTMs for semantically coherent lyrics generation.
- Conducted an extensive ablation study to evaluate performance of proposed model architecture, which proved to outperform baselines. Trained and tuned numerous variations of the model on AWS to improve scalability.

Sketch Generation using Generative Adversarial Network Mar 2019
Coursework of Deep Learning Course

- Implemented sheep sketch generation with GAN using LSTM as generator and discriminator as baseline.
- Improved with Bi-directional LSTMs, Wasserstein GAN and used pre-trained generator to warm-start the model.

Google Analytics Customer Revenue Prediction Nov - Dec 2018
Kaggle Competition, Project of Machine Learning Course

- Wrote Python scripts to analyze data features and implemented feature engineering and preprocessing.
- Implemented Ensembled method of XGBoost, Light GBM and CatBoost to improve result and ranked Top 15%.