Weixiao(Kevin) Sang

The University of Texas at Arlington | (682)-283-4886 | sangweixiao@163.com

EDUCATION

M.S. in Computer Science, The University of Texas at Arlington (UTA) **Aug. 2019 – Aug. 2020**
M.ENG. in Computer Technology, Beijing University of Posts and Telecommunications (BUPT) **Aug. 2018 – June 2021**
B.E. in Automation, Beijing University of Technology (BJUT) **Aug. 2013 – June 2017**

COURSEWORK

Computer Science
Foundation of Computational Linguistics, Artificial intelligence, Machine Learning, Data Mining, Algorithm, Probability Theory and Stochastic Process, Formal Languages and Automata, Data Warehouse & Data Mining, Distributed Computing Environment
Automation
Automatic Control Principle, Digital Electronic Technology, Analog Electronic Technology, Integrated circuit analysis and design, Embedded System

PERSONAL PROJECTS

Machine learning, Data mining, AI Project in UTA **Aug. 2019 – May 2020**
-Designed a **Naive Bayes** Classifier (without using scikit-learn), fully grasp this technique and test various smoothing methods.
-Implemented **Polynomial regression**, using scikit-learn to get weights and using root mean square error to calculate error.
-Implemented **KNN** algorithm with Euclidean distance, normalized Euclidean distance, cosine similarity in **Python** (without using any library), learned clustering method.
-Comment Rating system, using **Ensemble Methods** with naive Bayes method. Using **Python** and **Flask** and deploy to website.
-Implemented Decision tree, Bayesian network and Graph Search (without using any library).
-Text classification with **CNN**, **LSTM**, using Word2Vec and Glove **word-embedding**, finetuning **Bert** pretrain model to build text classification model.

Millions level Text classification (without using any machine learning library) **Aug. 2018 – Dec. 2020**
-Data mining course project in BUPT, I designed an optimized Naive Bayes classifier get 0.94 F1 score on **Millions level** Text of ten classes. Earned bonus points for working independently.
-Because the number of words in each class of text is very large, using **chi-square** to build word bag model and reduce word bag model dimension. Meanwhile, use **sparse matrix** to save text feature vector.
-Using Naive Bayes to build classifier, using Laplace add one smoothing. Using **TF-IDF** to modify feature vector.
-Using SVM Linear SVC and SGD as baseline, my model reaches the same F1 score as baseline, meanwhile, my model is faster than baseline (Linear SVC number of iteration is 100). Time for training and classify is 114s.

Relation classification with Attention biLSTM
-Using Glove word embedding, **bidirectional long short-term memory network** and **Attention** mechanism to build a relation classifier. Using SoftMax cross entropy as loss function. Using one-hot to build label.
-Dataset is SemEval2010 task8, 9 classes, the final F1 score is 0.765.
-Using python, Tensorflow, pytorch.

NER with CRF and biLSTM
-Implemented Named Entity Recognition (NER) Algorithm development.
-Using Glove word embedding, bidirectional long short-term memory network, linear **CRF** layer. Using 0-1 mask to build target. Using mean of CRF maximum likelihood as loss. Dataset is manual data. Final Precision is 0.97.

Seq2Seq translator with Attention
-Implemented Encoder-decoder.
-Using LSTM build encoder, LSTM and Attention to build decoder. Dataset is anki cmn-eng (Chinese-English). Using Glove word embedding, loss is sum of decoder output with NLLLoss.
-Using python, pytorch.

Text classification with LSTM and Self-attention
-Implemented Self-attention, self-attention can reduce long term forgetting.

Text classification with Multi-heads attention
-Implemented Multi-heads attention. Using this project to help me to understand Transformer better.

Chinese words segmentation
-Chinese words segmentation, using Bigram method, Laplace add one smooth method and dynamic programming. Time for segment the entire test set (100 sentences) is 0.08s.

User Management system
-A course project in Adv topics in database systems course in UT-Arlington. Designed a user information management system, user can upload csv file, and manage information in the system.
-Developed functions upload csv file, upload photo, delete user, add user, update data, search data, etc.
-Implemented back-end by using **SQLAlchemy** and **Flask, SQLite**.
-Deployed to IBM cloud.

Paper Note platform **Dec. 2019 – Nov. 2019**
-A course project in Software Engineering course in UT-Arlington. Designed a paper note sharing platform, people can read paper and write and share note on each section of paper, all the note can be commented.
-Defined two kind of user with authority including administrator and user.
-Developed functions sign in, sign up, edit profile, write note, write comments, upload paper, manage user, etc.
-Implemented front-end by using **Vue** and **Ionic** framework.
-Implemented back-end by using **SQLAlchemy** and **Flask**.
-My job is proposing functions and using Vue and Ionic framework to build front-end (note page, sign in, sign up, profile page).

ServeMe Platform **Feb. 2020 – Mar. 2020**
-A course project in Adv Tops Software Engineering course in UT-Arlington. Designed a Home Services system, have making new request, bidding and order management functions.
-Two kinds of user, customer (search and place service request) and services provider (place bid for request and provide service).
-Using **Ionic** and **Angular** as front-end framework implement android client.
-Implemented back-end by using **SQLAlchemy** and **Flask**.
-Follow the scrum method (agile development spirit). Have regular standing meetings and finish each sprint for two weeks.
-My job is mainly using ionic framework and design the front-end (user information part, include HTTP request and service).

Design of micro spectrometer based on MCU and CPLD **Dec. 2016 - May 2017**
-A Undergraduate Graduation Design in Beijing University of Technology
-Using Verilog to design a CPLD control program, Using C++ to design an MCU control program, Using QT-5 to design human-computer interaction interface.

SKILLS

Language: Python, C language, C++, SQL, Java
Tools: Pycharm, Git, Jupyter Notebook, Mysql, SQLyog, IBM cloud, Google cloud
Frameworks: Pytorch, Tensorflow, Ionic, Vue, Angular, Flask
Platform: Windows, Ubuntu

AWARDS AND PUBLICATION

Awards
First-class scholarship in BUPT **October,2018**
First-class scholarship in BUPT **October,2019**

Publication
Physics and Engineering 2014-2（CSTPCD）-The experimental error analysis of gravity acceleration by using free falling