SHIXIANG WANG

E-mail : Shi-xiangWANG@outlook.com · Phone : +86-181-3797-2556 · Birthday : 1997.02

EDUCATION

CityU, Hong Kong **MSc in Computer Science** **GPA: 3.75 / 4.30** **2021.08 – Now**

Course: Machine Learning A&B, Nature Language Processing, Data Mining and Data WareHousing, Artificial Intelligence

POSTECH, Korea **Department of Mathematics** **Exchange Student** **2019.08 – 2019.12**

Course: Data Science, Actuarial Mathematics, English Writing

HIT, China **BSc in Information and Computing** **GPA: 86.4 / 100** **2016.09 – 2020.06**

Course: Probability and Statistics(96), Numerical Methods of Partial Differential Equation(96), Numerical Analysis(92), Operations Research(100), Equations of Mathematical Physics(91), Elements of Information Science(96), Preliminary Functional Analysis(98), Function of Complex Variable(97), Topology(93), Differential Geometry(100), Discrete Mathematics(84), C Language(90), Data Structure(81), Computer Network Technology and Applications(90)

RESEARCH EXPERIENCE

An Improved Gradient-based Meta-Learning algorithm for Large Scale Drug Discovery

- Applied an innovative 'MetaDrug' algorithm on a new large-scale, high-quality drug discovery benchmark data obtained from the ChEMBL database using the RDKit package
- Compare with the MSE of the standard training process and the baseline model(Random forest), the average MSE of 'MetaDrug' is reduced by at least 80% after one gradient descent, and the error convergence speed is faster. The mean and median R^2 values over all meta-testing assays improve a lot on four different testing groups compared to MAML

Question Generation: Finetuned on Pretrained with QA-pair Evaluation

- Finetune BART and T5 models on the preprocessed SQuAD dataset. Then innovative use BERT to evaluate the generated questions and provide answer pairs to choose the most matched question
- Ensemble BERT-Evaluator greatly improves the BLUE, ROUGE and METEOR score. Report can be found *here*

PROJECT EXPERIENCE

Kaggle: Animal cuteness analysis 2021.11

Regression task: Use the provided raw data and raw pictures to predict the cuteness value of animals

- Perform linear regression, ridge regression, LASSO, random forest regression on the provided original features, and use PCA to extract core features. Extract features directly from the original image and apply deep learning models such as CNN. Transfer learning with existing models such as EfficientNetB7
- The minimum RMSE after transfer learning of EfficientNetB7 is 20. Codes can be download *here*

Kaggle: Tweet Sentiment Analysis 2021.09

Classification task: Sentiment category judgment and analysis of tweets about Google, Apple, Twitter, and Microsoft

- Analyze tweet features, remove special symbols, common words, and lemmatization. Use BOW and TF-IDF to extract dictionaries from the cleaned data and vectorize the training text.
- Use Naive Bayes, Logistic regression, SVM, Adaboost, etc. to train separately, and use grid search to select the optimal value of hyperparameters. Codes can be download *here*

INTERN AND WORK EXPERIENCE

Cambricon, Beijing **Algorithm Research Internship** **2022.05 – Now**

- Research field: Monocular 3D Object Detection
- Read papers about monocular detection such as SMOKE, FCOS3D, PGD, DD3D
- Reproduce non-open source code such as PGD. Use mmdet3d to implement DD3D
- Design experiments to improve the performance of the above algorithms

JD.com, Beijing **Software Development Engineer** **2020.09 – 2021.07**

- Techniques: Jsf, Mq, Redis, Spring, Mybatis, Git
- JDL order receiving system requirements development (agile development)
- Over 20,000 lines of code during work
- Excellent employee of the team in 2020

AWARDS

Renmin Scholarship $\times 5$, Outstanding Student, Provincial First Prize in College Students Physics Competition

SKILLS

- **Program Language**: Familiar with Python, Pytorch, and Java
- **English Ability**: IELTS 7.0 （Listening 7.0, Reading 8.0, Writing 7.0, Speaking 6.5）
- Solid foundation in data structures and machine learning. Reproduce LS, LASSO, RLS, K-means, EM-GMM, and Mean-shift algorithms
- Familiar with RNN, LSTM, Seq2Seq, Attention, Transformer , and Bert algorithms
- Familiar with FCOS3D, PGD, SMOKE, and other detection algorithms in Open-mmlab
- High self-motivated and passion