SHENGJIA YAN

(+86)15961887272 · sjyan1995@gmail.com · yanshengjia.com · https://github.com/yanshengjia

Education

- **Southeast University (SEU)** 2013.08 - 2017.06
  B.E. in Computer Science; GPA: 3.56/4.0 *Nanjing, China*

Working Experience

- **17zuoye AI Research Group** 2017.06 - Present
  NLP Engineer *Beijing, China*
  - Led a team of six to design and develop an automated essay enhancing system
  - Built an automated essay score predictor based on Feature Engineering, Logistic Regression and LSTM, which reached best performance on Kaggle ASAP dataset.
  - Developed a grammar checker based on Convolutional Seq2Seq and rules, which reached best F0.5-score on CoNLL2014 dataset. Reduced deep learning inference time by 50% by utilizing Nvidia TensorRT.
- **Knowledge Science and Engineering Lab @ Southeast University** 2014.10 - 2017.06
  Research Assistant (advisor: Prof. Guilin Qi) *Nanjing, China*
  - Conducted data preprocessing using NLP approaches like spaCy to refine and analyze datasets.
  - Presented and implemented a Random Walk algorithm in Python based on Probabilistic Graphical Model to map the string mentions in web tables to their referent entities in a knowledge base.
  - Achieved 6% increase in F1-score compared with the state-of-the-art scheme. The result was **published in [1, 2]**.

Selected Projects

- **Deep Learning Grammar Error Correction System** 2018.06
  - Designed and developed a GEC system based on Facebook well-known Convolutional Seq2Seq paper and rule-based proofreading software LanguageTool.
  - Supported grammar checking of more than 1000 tokens per second on single Tesla P100 GPU by optimizing deep learning inference with Nvidia TensorRT and ONNX.
- **Crowdsourcing NLP Annotation Platform** 2018.05
  - Designed a crowdsourcing annotation system with multiple quality control mechanisms based on annotation tool BRAT.
  - Developed the frontend with HTML, Bootstrap and Javascript.
  - Built the backend service using Tornado/Python, MongoDB and deployed on AWS.
- **DNN-Based Face Recognition System** 2017.03
  - Implemented the neural network Backpropagation algorithm in C and Constructed a DNN to recognize human's face, pose and experssion.
  - 145+ stars and 180+ forks on GitHub
- **C-Minus Compiler** 2016.06
  - Implemented the Regular-Expression-to-NFA converter, LR(1) parser and semantic analysis module in Python.
  - Visualized the compiling process by ploting NFA, DFA, GOTO graphs with GraphViz.

Publications

1. "Entity Linking in Web Tables with Multiple Linked Knowledge Bases", In: *Semantic Technology: 6th Joint International Conference: JIST 2017*. pp. 239-253 [pdf]
2. "A Method of Entity Linking in Web Tables based on Multiple Linked Knowledge Bases", Chinese Patent, CN106503148A, 2017

Skills

- Languages: Python, C/C++, JavaScript, Markdown, LaTeX
- Tools: Git, MongoDB, Tornado, Bootstrap, Qt
- Frameworks: TensorFlow, PyTorch, Scikit-Learn, Gensim, spaCy

Honors

- Computer Programming Contest (Jiangsu Province), Third Prize, 2016.11
- SEU Computer Programming Contest, Fourth Place, 2016.10
- Outstanding Project, SEU Student Research Training Program, 2016.05

Extracurricular Activities

- International Student Leadership Program, California Polytechnic State University, CA, USA, 2016.01
- Nanjing Youth Olympic Games Volunteers, National Olympic Committee Assistant, Nanjing, China, 2014.07