Bojun Jin

Email: 200110828@stu.hit.edu.cn
Phone/WeChat: +86 18865038551

Basic Information

- Grade: 92.013
- Ranking: Top 9%
- CET-4: 634
- CET-6: 547
- Major: Computer Science
- School: Harbin Institute of Technology, Shenzhen

Research Experience

HITSZ-HLT

- 2022 AI-Debater Competition: Responsible for the Interactive Argument Pair Identification task. Complete data preprocessing, k-fold cross-validation and Ensemble Learning based on bert-large, RoBERTa, ALBERT, ERNIE, RoFormer.
- Baidu PaddlePaddle Learning Race: Based on pre-training + fine-tuning, we implemented the NER and Sentiment Analysis tasks and achieved great scores. To solve the problem of uneven label distribution, we use Label Smoothing and Unbalanced Loss Function. Furthermore, We used SMART adversarial training to solve the problem of overfitting.
- 2023 AI-Debater Competition: Responsible for the task of argument generation. Use ChatGPT and the strategy of combining short argument points for data augmentation. By comparing the effect of Bart, T5 and CPT models, as well as the performance of several decoding strategies, the best combination is selected.
- Context Selection Strategies: When predicting the stance of one argumentative component to its parent component in a tree structured dataset, suitable context component information is helpful to the performance. Therefore, I investigated some context component selection strategies to find which strategy achieves the best performance.
- Investigated Argument Mining task including component segmentation, component classification, relation detection and relation classification. The main methods include generative models and many kinds of discriminative models.

Research: Argument Quality

The main work of the research is to predict which argument in a argument pair is more convincing to the topic.

- The relative position of the two arguments has an important effect on the prediction of the model. By exchanging the relative position of the two arguments for training, the prediction performance has been greatly improved.
- Predicting the good or bad quality label and the quality score of a single argument at the same time, and adding their losses for training, has effectively improved the accuracy of prediction.

Project

RISC-V CPU

- Based on Verilog HDL, implement a single period CPU and a five-stage pipeline CPU which brings about five-fold increase in CPU efficiency. All the modules(including logic gates) and circuits are implemented by Verilog HDL on Vivado.
- Particularly, due to the introduce of forward module and branch prediction module, the pause time is significantly reduced and the CPU efficiency has been further improved.
- Completed all optional CPU instructions. The whole process of the project went smoothly and scored 97 points.

Campus Canteen Food Delivery Ordering System

- Based on Java and MySQL database, implement a campus canteen delivery food ordering system with a UI front-end and a MySQL database back-end. There are three types of users(customer, merchant, canteen administrator), each with their own functions.
- Fully functional, including but not limited to registration, login, menu browsing, customer ordering, order status update, menu maintenance, canteen management, etc.
- Design a database schema for this system, including tables, triggers, constraints, etc.

xv6 OS Kernel Optimization

Complete all the experiments required by our school's OS experiment course, the following are some representative results:

- Add system call trace: It can output the system call sequence of a process which is helpful for debugging.
- Parallelization based on fine-grained locks: Optimize the xv6 memory allocator for multi-core scenarios, change from all CPUs competing for one free list to each CPU maintaining a free list and obtains physical page frames from its own free list first, which significantly reduces lock contention.

Computer Network Protocol Cluster Experiment

Using C language to implement the network protocol cluster, including the function of the real protocol cluster for frame reception, layer by layer unpacking, layer by layer packaging, package sending.

- Independently designed and implemented various common protocols of computer network, including UDP in the transport layer, IP and ICMP in the network layer, Ethernet in the data link layer, and the address resolution protocol ARP.
- Based on the key-value pair container that supports the entry validity period, it simplifies the implementation of ARP caching and other functions.
- In this protocol cluster experiment, different layers of protocols are decoupled through modular design, which is conducive to improvement and modular development.

Awards and Honors

- Freshman and Sophomore Annual Scholarship of Harbin Institute of Techonology, Shenzhen
- 2022 AI-Debater 1st Prize in Track3
- Excellent Communist Youth League Member of Harbin Institute of Technology, Shenzhen

Skills

- Linux Command
- Programming Language: C/C++, Java, Python, MySQL
- Program Tool Software: VS Code, MySQL, CodeBlocks, Vivado, IntelliJ IDEA, Android Studio, RARS Assembler Simulator