Hejie Cui

https://github.com/HennyJie

Education

Emory University   **Atlanta, GA, United States**
PhD in Computer Science & Informatics   *Aug.2019-currently*

- **Courses:** Artificial Intelligence; Machine Learning; Algorithms; Data Mining; Information Retrieval

Tongji University   **Shanghai, China**
B.Eeng. in Software Engineering   *Sep.2015–Jul.2019*

- **GPA:** 4.9/5.0 (rank 1st/164 or top 0.6%)
- **Awards:** 1) National Scholarships (awarded to top 0.2% nationwide) for 3 consecutive academic years; 2) 2019 Outstanding graduates of Shanghai, China (top 0.1%); 3) Honored Student of Tongji University (top 2.5%)

Skills

- **Programming Languages:** Adept in Python, Experience in C++, R, MATLAB, JavaScript
- **Development:** iOS, Web
- **Tools and Platforms:** Git, Jetbrain, Linux, PyTorch, MySQL

Research Experience

COVID-19 Search   **Atlanta, USA**
CSI @ Emory University, directed by Prof.Eugene Agichtein   *May.2020–Currently*

- Target: given a topic related to COVID-19, produce a ranked list of documents from a collection of biomedical literature articles per topic ordered by decreasing likelihood that the document matches the information need
- Participate in round 2 with a two stage method: Stage 1.Baseline Replication from Anserini; Stage 2.Rerank with SciBERT
- Add published time as new features for reranking
- Still going on and plan on involving more social signals (tweets/retweets/likes and arxiv downloads) as additional features in round 3

Visual Question Answering   **Atlanta, USA**
CSI @ Emory University, directed by Prof.Eugene Agichtein   *Jan.2020–May.2020*

- Based on the work of Deep Modular Co-Attention Networks, add a module of OCR type question classification
- Extract attention of ROI for OCR: for each image, extract a distribution showing the attention of different ROIs from model, in order to find the region with the highest probability to produce correct answer
- Add a self-attention to enhance fused feature

Mining of Potential Influencing Factors for COVID19 Spread   **Atlanta, USA**
CSI @ Emory University, CS570 Data Mining   *Mar.2020–May.2020*

- Use GAM model to check whether environment factors have influence on spread of coronavirus
- Use SIR model to check whether non-pharmaceutical intervention can help to prevent the spread of coronaviru
- Fit a Recurrent Neural Network (RNN) model to predict the daily new confirmed cases of tomorrow based on government response and the daily new confirmed cases of today

Automatic Commit Messages Generation from Diffs   **Atlanta, USA**
CSI @ Emory University, CS557 Artificial Intelligence   *Oct.2019–Dec.2019*

- A hybrid method: TF-IDF ranking method improved with a Seq2Seq model based on pointer generator network
- TF-IDF part: given a diff string, find the most similar diff and get its comment as candidates
- Generate model part: use pointer generator network to predict the next word in the target sequence, re-rank the top 10 matching results from the IR method by the possibility matrix obtained from Seq2Seq model.

DCOL for Nonlinear Distance Calculation Applied to PCA and T-SNE   **Shanghai, China**
[Remote] Rollins School @ Emory University, directed by Prof.Tianwei Yu   *Nov.2017–Feb.2018*

- Proposed a new kernel dimension reduction method based on DCOL(Distance Based on Conditional Ordered List), which could reveal strong nonlinear dependencies in the data
- Adopted squares instead of absolute values and made a transformation on DCOL matrix to make the kernel have a new property
- Compared the dimension reduction result of the new method with kernel PCA, PCA and T-SNE, and found that the information consistence was increased by introducing the new non-linear distance

Detection and Distance Measurement of Speed Bumps   **Shanghai, China**
Computer Vision Lab @ Tongji University, directed by Prof.Lin Zhang   *May 2018–Jul.2018*

- Developed an integral speeds bumps detection and distance measurement system for no-man sweeper vehicles by using Python programming
- Utilized Yolo v3 net to detect the speed bumps and obtain the position and size of bounding box
- Detected the speed bumps in the video and outputted the distance between the detected bump and the bracket in real time, optimized the model by redefining the distance calculation

Anomaly Detection Framework using Machine Learning Methods   **Shanghai, China**
AI Healthcare Lab @ Tongji University, directed by Prof.Jianwei Lu   *Mar.2018–Jun.2018*

- Established a new framework for anomaly (CPU, Memory, IO) detection and stress testing, which can forecast potential failures and pressure spills based on performance data
- Collected random injection failure and normal data by using Clear Water platform, trained the classical set KDDCUP99 and data collected in true environment through machine learning classifiers (SVM, Random Forest, NN, etc.)
- Contrasted the precision, recall rates and F1 score of different classifiers, disovered the highest accuracy (99.7%) with using NN methods

Internship Experience

SENSETIME   **Beijing, China**
Beijing Research Institute   *Dec.2018-Jul.2019*

- Worked as a intern algorithm engineer in the Intelligent Medical Group
- Developed a pulmonary vessel segmentation algorithm based on my updated network, an orthogonal fused U-Net++, for chest CT images
- Published a patent on my intern work and got one paper accepted by MICCAI (International Conference on Medical Image Computing and Computer Assisted Intervention, which is the tier 1 conference in medical imaging field) 2019 as the first author

Improve Center Line Tutor by Deep Learning   **Ontario, Canada**
Perk Lab @ Queen's University, directed by Prof.Gabor Fichtinger   *Jul.2018–Oct.2018*

- Built an extension for classifying web-cam video images using Tensorflow in 3D Slicer
- Used Tensorflow in real-time workflow detection for providing real-time feedback in central venous catheterization training
- Made distortion such as deforming, cropping, or brightening in the training inputs in random ways to polish the model, analyzed the influence of each parameters to get the best retrained model

SAP   **Shanghai, China**
Shanghai Engineering Labs   *Jun.2017-Aug.2017*

- Helped to develop SAP ERP system and use the HANA database to process enterprise management data.

Leadership and Activities

- President of Tongji University Microsoft Student Club
- Undergraduate Assistant at iLab, Tongji University
- 2018 CSC Full Scholarship for MITACS Globalink Research Internship in Canada
- 2017 Tongji University Programming Competition (Second Prize)
- 2017 American Mathematical Contest in Modeling (Meritorious Winner)
- 2017 Best Work for Microsoft Summer Camp Hackthon

Publication

- **Hejie Cui**, Xinglong Liu, Ning Huang, Pulmonary Vessel Segmentation based on Orthogonal Fused U-Net++ of Chest CT Images, MICCAI (Medical Image Computing and Computer Assisted Intervention) 2019.
- Huijuan Zhang, Chengxin Jin, **Hejie Cui** (2018) A Method to Predict the Performance and Storage of Executing Contract for Ethereum Consortium-Blockchain, ICBC2018 P.63-74.

Teaching

- **CS584 Deep Learning in BioMedical**, Teaching Assistant, Spring 2020.
  Course Website: https://reynalab.github.io/courses/cs584_spring2020.html