Jie Liu

328 North Craig Street – Pittsburgh – PA, 15213

+1 469-910-4827 • jil251@pitt.edu • github.com/caterby

RESEARCH INTEREST

- Machine Learning, Big Data Management and Analysis, Convex Optimization, Graphic Models.
- Apply Machine Learning techniques to solve real-word problems, like Medical Images, Bioinformatics, Data Security, Anomaly Detection, etc.

EDUCATION

The University of Texas at Dallas, Richardson, TX Aug. 2016 - May. 2018
M.S. in Computer Science *GPA: 3.70/4.00*

Xian Jiaotong University, Shaanxi, China Aug. 2013 - May. 2016
M.S. in Software Engineering *GPA: 81.05/100*

Northwestern Polytechnic University, Shaanxi, China Aug 2009 - May 2013
B.S. in Software Engineering *GPA: 3.52/4.00*

PROFESSIONAL EXPERIENCE

Graduate Research Assistant - Data Management and Analytics Lab
University of Texas at Dallas, Richardson, Texas *Mar. 2017 - May. 2018*

- Developed efficient techniques for addressing challenges in Data Stream Mining, e.g., Concept Drift (evolving class boundary), emergence of novel classes, scarcity of labeled data, Sampling-bias, Class Imbalance, etc.
- Improved scalability by proposing parallel and distributed algorithms using Big Data analytic tools.
- Applied developed techniques on data mining applications, including security domain, e.g., anomaly detection.

Intern Fellow - Detect Game Cheating Team
NetEase Games, Beijing, China *May. 2018 - Aug. 2018*

- Collected and cleaned data set out of noisy real-world raw data for analysis using various data pre-processing steps.
- Developed a transfer learning approach to detect cheatings of computer games for server-side cheat detection.
- Implemented and deployed our proposed approach in distributed environment through Spark for scalability.

Teaching Assistant - Computer Science Department
University of Texas at Dallas, Richardson, Texas *Aug. 2016 - Jan. 2017*

- Assisted in teaching of several graduate and undergraduate level courses, e.g., Algorithms Design, Machine Learning.
- Helped students to understand basics, resolve issues, arrange demo for projects, and grade related parts of the course.

RESEARCH PROJECTS

Multistream Classification with Relative Density Ratio
(Data Mining, Python), CSE Dept, UT-Dallas *Oct. 2017 - Mar. 2018*

- Completed a multi-stream classification framework with asynchronous concept drift detection and covariate shift adaption.
- Proposed a gradient-based technique to automatically learn model parameters for relative density ratio estimation from available data.
- Used real-world and synthetic datasets to empirically evaluate our approach, and compared the results with baseline.

Research and Application of Word and Text Similarity Computing Method
(Text Data Processing, Text Mining , Java), XJTU *Aug. 2015 - Jun. 2016*

- Designed and implemented a novel algorithm for computing the similarity of Chinese words which merges synonyms based on the semantic information provided by a most widely used Chinese Thesaurus.
- Developed algorithms for computing the Chinese text syntactic and semantic similarity via maximum bipartite matching.
- Built a system named Abstract Similarity Comparison through various text mining techniques like text parsing, TF-IDF, Vector Space Model, which can be applied into Text Mining and Plagiarism Detection.

Multistream Regression with Direct Density Ratio Estimation

(Data Mining, Online Learning, Python), CSE Dept, UT-Dallas *Apr. 2017 - Oct. 2017*

- Proposed an online regression model which utilizes an efficient technique that simultaneously performs bias correction and asynchronous concept drift detection between source and target stream data.
- Analyzed the theoretical properties of our approach to show its effectiveness in addressing the challenges of multistream regression setting.
- Experiment results on benchmark data sets indicate significantly improved performance over the existing methods.

ACADEMIC PROJECTS

Time Serial Data Visualization for American Airline Companies' evolvement

- **Tools and Technology:** JavaScript, Matlab
- Prepared dataset out of real-world raw data and reduced the data dimension via Principal Component Analysis(PCA).
- Designed a comprehensive framework for high-dimensional data visualization via JavaScript and Clustering technique.
- Users can observe the Airline Companies' evolving process through this visualization model.

Design and Implementation of Personal Management System

- **Tools and Technology:** ASP .NET, UML, MVC, MS SQL, HTML, CSS
- Designed the main framework of this system, and completed the model for software system through UML..
- Implemented the core function of this system, and solved variety of user problems including account set-up, password issues, database design, software configuration.
- Customers can browse categories, search detail info about any person through key words, Admin can alter, add, remove person info, setup special promotion etc.

Brain Tumor Segmentation

- **Tools and Technology:** Python, Matlab, Tensorflow.
- Set up a model that can process large amount of medical image data in reasonable time.
- Extracted the image features using several approaches such as Local Binary Pattern (LBP), Scale-invariant Feature Transform(SIFT), Features from Accelerated Segment Test(FAST).
- Classified images into sub-categories via Softmax and CNN, analyzed their performance based on different data size.

PUBLICATIONS

- Bo Dong; **Jie Liu**; Swarup Chandra; Latifur Khan, "Multistream Classification Using Relative Density Ratio Estimation", *To Be Appeared: 33th AAAI Conference on Artificial Intelligence, Hawaii, USA, 2019*
- Haque A.; Hemeng T.; Chandra S.; **Jie Liu**; Khan L., "A Framework for Multistream Regression with Direct Density Ratio Estimation", *In: 32th AAAI Conference on Artificial Intelligence, New Orleans, Lousiana, 2018*
- Chandra S.; Haque A.; Hemeng T.; **Jie Liu**; Khan L., "Ensemble Direct Density Ratio Estimation for Multistream Classification", *In: 34th IEEE International Conference on Data Engineering, Paris, France, 2018*

TECHNICAL SKILLS

- **Programming Languages:** Java, Python, Scala, C, C++, ASP.NET, HTML, CSS, JavaScript
- **Database & Visualization:** MS SQL, Teradata, Oracle, MySQL, Tableau, PL/SQL, LINQ
- **Data Analysis:** R, Weka, Scikit-Learn, NLTK, Hadoop, MATLAB, Tensorflow

RELEVANT COURSEWORK

- Data Visualization(A) Artificial Intelligence(A) Design and Analysis of Computer Algorithms(A)
- Machine Learning(A) Natural Language Processing(A-) Statistical Methods in AI and ML(A)

LANGUAGE

- **GRE: 324 + 3.0** Verbal: 156; Quantitative: 168; AW: 3.0
- **TOEFL: Waive**