Zhou Jiehui (周杰辉)

Github: github.com/algzjh

Personal Page: algzjh.github.io

Email: jiehuizhou1997@gmail.com

Mobile/WeChat: 18932449056

I am currently pursuing my Ph.D. degree at the State Key Laboratory of CAD&CG, School of Computer Science and Technology, Zhejiang University, China, under the supervision of Prof. Wei Chen. My research topic is heterogeneous data mining and visual analytics in healthcare, specifically involving **visual analytics**, **causal analysis**, **out-of-distribution detection**, and **privacy preservation**. I am looking for a job in big data analytics with a focus on the following areas: data visualization and trustworthy artificial intelligence, such as explainable AI and causal inference.

EDUCATION

- **Central South University** — Changsha, China
  Bachelor of Computer Science & Technology; GPA: 91.53/100 — *Sep 2015 - June 2019*

  Courses: *Data Structures, Algorithm Analysis and Design, Operating Systems, Databases, Computer Networks, Probability Theory and Mathematical Statistics, Machine Learning, etc.*

- **Zhejiang University** — Hangzhou, China
  PhD of Design Science; State Key Lab of CAD&CG; GPA: 88.49/100 — *Sep 2019 - June 2024 (plan)*

  Courses: *Frontier Technology of Big Data Visualization, Big Data Storage and Processing, 3D Animation and Interactive Technology, Cloud Computing Principles and Application Development, etc.*

PROJECTS & PUBLICATIONS

- **Spatio-temporal Correlation Analysis System of Health Insurance Data (Multi-dimensional Correlation Analysis, Spatio-temporal Data Visualization**): Application-oriented. *Problem*: Multiple types of health insurance fraud behaviors are intertwined in dimensions such as participants, time and space, making analysis difficult. *Approach*: (1) Design frequent pattern mining and anomaly detection models to quickly screen multiple types of suspicious behaviors, such as multi-location prescribing and false hospitalization; (2) Develop a visual analytics system with multi-view likage, providing time series and geographic visualization. Users can be supported to interactively identify, explore, and verify associated frauds. *Evaluation*: The system was actually deployed in the Health Insurance Bureau of Fuyang District to help detect participants and doctors involved in various types of fraudulent behaviors, and received positive feedback from users.
  Publication: Zhou Jiehui, Zhu R, et al. MedicareVis: a Joint Visual Analytics Approach for Anti-Fraud in Medical Insurance. Journal of CAD & CG, 2021. (EI) [Paper]

- **Visual Analysis of Medical Collusive Fraud (Multi-source distributed Data, Graph Modeling, Anomaly Detection, Human-Computer Collaboration**): Application-oriented. *Problem*: Medical collusive fraud involves multiple patients and locations, and is difficult to distinguish from normal patient behavior (e.g., regular follow-up visits for chronic disease groups), leading to high false-positive detection rates and requiring significant time for manual verification. *Approach*: (1) Considering the spatio-temporal association characteristics of collusive fraud, we construct a weighted network to portray the common visit behavior among fraudsters, and then use the community detection model to mine the suspected collusive fraud groups. (2) Propose similarity measures between patients based on diseases and drugs to help users determine the rationality of patients' multiple visit. (3) Develop an interactive system, including visual designs such as graphs and event sequences, to support users to efficiently compare, explore, and verify suspected collusive fraud. *Evaluation*: Based on data from 2020-2021 in the Fuyang district, the system identified a total of 435 people with collusive fraud, involving 66 institutions and a total medical treatment amount of over ¥500,000.
  Publication: Zhou Jiehui, Wang X, et al. FraudAuditor: A Visual Analytics Approach for Collusive Fraud in Health Insurance. IEEE Transactions on Visualization and Computer Graphics, 2023. (CCF-A) [Paper] [Slide] [Demo]

- **Privacy-preserving Publishing Approach for Visualization Charts (Cross-domain Data Security, Differential Privacy, Bayesian Network, Human-Computer Interaction**): Research-oriented. *Problem*: Publishing visualization charts faces a trade-off between privacy breaches and insight dissemination. *Approach*: (1) Propose pattern constraints to model user preferences for different data patterns, and use them to guide a Bayesian network-based differential privacy-preserving data generation model; (2) Develop a visualization system to allow users to interactively select patterns and specify their importance, understand the privacy-preserving process of the model, and compare different privacy-preserving schemes. *Evaluation*: Quantitative experiments based on pattern utility metrics such as clustering, correlation, and ranking

demonstrate that the proposed approach significantly outperforms the existing privacy-preserving data generation model baseline at moderate levels of protection (significance level p-value < 0.01, medium effect size Cohen's d $= 0.5$).

Publication: Zhou Jiehui, Wang X, et al. DPVisCreator: Incorporating Pattern Constraints to Privacy-preserving Visualizations via Differential Privacy. IEEE Visualization Conference, 2022. (CCF-A) [Paper] [Code] [Demo]

- **Visual Analysis of Causal Heterogeneity (Subset Distribution Heterogeneity, Interpretable Clustering, Causal Inference, Human-computer Interaction)**: Research-oriented. *Problem*: In scenarios with heterogeneous data distributions, treatment are difficult to be predefined, which forces data analysts to repeatedly attempt treatment hypothesis, subgroup analysis, and effect validation. *Approach*: (1) Introduce concepts to describe data units with similar values of specific attributes to improve interpretability and flexibility; (2) Propose a concept-based model, HetMatch, to quickly identify heterogeneous contexts with potentially effective causal relationships by matching decision rules; (3) Develop a system with coordinated visualizations to support users to interactively edit concepts, compare causal relationships between subgroups and interpret treatment effects in terms of matched units. *Evaluation*: Quantitative experiments on simulated data demonstrate that HetMatch has a smaller effect estimation error than baseline method CausalTree for more than 10 variables. Case studies of real-world scenarios such as medical diagnosis and smartphone fault localization illustrate the usability of the system.

  Publication: Zhou Jiehui, Wang X, et al. CausalPrism: A Visual Analytics Approach for Concept-based Causal Heterogeneity Exploration. (under review)

EXPERIENCE

- **Alibaba** — Onsite
  Engineering Intern — *June 2020 - Aug 2020*
  - **Project**: Relationship mining and analysis of social data of Taobao and Xianyu users to assist in friend and product recommendation.
  - **Keywords**: Distributed Computing, Social Network Analysis.

- **Zhejiang Lab** — Onsite
  Research Intern — *Sep 2020 - May 2021*
  - **Project**: A visual analysis system for massive dynamic graph data, supporting the analysis of temporal associations and evolution patterns between different nodes.
  - **Keywords**: Data Fusion, Dynamic Graphs, Visual Analysis Systems.

- **OPPO** — Onsite
  Research Intern — *Sep 2022 - Jan 2023*
  - **(1) Project**: Root cause analyzing of faults based on the event tracking data of smartphones, mining causal relationships and effects between monitoring variables, and developing a system to explore and verify causal influence paths.
  - **(1) Keywords**: Cause Discovery, Root Cause Path Mining, Causality Interpretability.
  - **(2) Project**: Evaluate the performance of machine learning models under different usage scenarios. Using self-supervision to learn the representation of event tracking data, and then labeling a small amount of data based on customer feedback, using semi-supervised fine-tuning to obtain a rating prediction model. Finally, the contribution of each feature to the model rating is obtained using model interpretable methods such as SHAP.
  - **(2) Keywords**: Self-supervised Learning, Semi-supervised Learning, Multi-dimensional Tabular Data, Model Interpretable Methods.

SKILLS AND HONORS

- **Languages**: Python, JavaScript, C++, Pytorch, Scikit-learn, Django, React
- **Tools**: Git, MySQL, Docker, Linux Server
- **Soft Skills**: Writing, Project Management, Time Management, Public Speaking
- **Community Activities**: Zhihu (Trustworthy Artificial Intelligence), Yuque (Causal Inference, How to Find Papers, How to Read&Write a Visualization Paper)
- **Awards**: National Scholarship (2016), Silver Medals of ACM-ICPC Asia Regional Competition in Nanjing and Xuzhou (2018), Outstanding Graduate Student (2021)