Haoyuan Zhang

CONTACT INFORMATION	CS335, Peter Landin Building, Queen Mary University of London, London E1 4NS, United Kingdom.	*Tele:* +44 (0) 7719232747 *Email:* haoyuan.zhang@qmul.ac.uk *Homepage:* haoyuan.uk

EDUCATION

Queen Mary University of London (QMUL) — London, UK
Ph.D. in Computer Science — Sep. 2015 - Jul. 2019
- Thesis: *A Bayesian-Based Framework for Making Inspection and Maintenance Decisions from Data and Expert Knowledge*
- Supervisors: Dr. William Marsh, Prof. Norman Fenton, Prof. Martin Neil

The University of Hong Kong (HKU) — Hong Kong
M.Sc. in Industrial Engineering and Logistic Management — Sep. 2013 - Dec. 2014
- Thesis: *Colour Petri Net-Based Modelling for Integrated Process Planning and Scheduling* (obtained the highest grade among the department)
- Supervisor: Dr. Tak Nam, Wong

Jinan University (JNU) — Guangzhou, China
B.Mgt. in Electronic Commerce — Sep. 2009 - Jul. 2013
- Thesis: *Tourism Supply Chain Collaborative Demand Forecasting Model Based on Colour Petri Net* (awarded the best undergraduate thesis of Jinan University)
- Supervisor: Dr. Bai Hua

RESEARCH EXPERIENCE

Project: Knowledge Discovery from Health Use Data (KNIFE) (with NHS Clinical Commission Group) — London, UK
Pilot Research Project - Funded by Alan Turing Institute — Apr. 2019 - Aug. 2019

- *Data Processing* - Use PostgreSQL to select cohorts of patients from clinical databases and preprocesses selected cohorts (e.g. missing data imputation).

- *Data Analysis* - Perform Exploratory Data Analysis (EDA), including feature projection (e.g. Principal Component Analysis (PCA)), feature selection (e.g. random forest), Conditional Independence (CI) tests (e.g. mutual information).

- *Model Building and Validation* - Build Bayesian Networks based on elicited medical knowledge and CI tests, fit their parameters with Maximum Likelihood estimates or Bayesian Parameter estimates, and compare the fitted models' performance with other machine learning techniques.

- *Application* - a) Generate synthesis clinical data from the learned models using forward/rejection sampling to bypass the restriction of health data usage; b) Use Structural Equation Modelling (SEM) to study the causal relationships between various latent variables.

Project: Player Pathways: Understanding Career Paths That Deliver Success for Professional Football Players and Clubs (with PlayerLens) — London, UK
Data Study Group - Funded by Alan Turing Institute — Dec. 2018

- *Feature Selection* - Used random forest to help club owners understand which features are the most influential features on players' values.

- *SEM and Network Analysis* - Worked with team members to perform SEM and network analysis on the selected features to form causal relationships.

Project: Predicting Language Outcome and Recovery After Stroke (PLORAS) (with University College London) London, UK

Data Study Group - Funded by Alan Turing Institute May. 2018

- *Feature Engineering* - Led the theme of feature engineering, which includes random forest for feature selection and PCA for feature projection. These features were used by other team members for statistical analysis and recovery prediction.

- *Survival Analysis* - Performed survival analysis to estimate the probability of whether a patient will recover given a time.

Workshop: The Nature of Questions Arising in Court That Can Be Addressed via Probability and Statistical Methods Cambridge, UK

Workshop - Funded by Isaac Newton Institute, University of Cambridge Aug. 2016

- Discussed applications of statistical methods and probabilistic reasoning in forensic science.

PROFESSIONAL EXPERIENCE

Queen Mary University of London London, UK

Research Assistant for the KNIFE Project Apr. 2019 - Aug. 2019

- Work on medical data mining, which includes causal inference, medical diagnosis, patient condition prediction and data generation. This project aims to deliver a proposal for grant application.

Teaching Assistant Sep. 2016 - Apr. 2019

- Lab/tutorial demonstration and coursework marking (average 50+ students per module) for:
  * ECS647U/ECS773P - Bayesian Decision and Risk Analysis,
  * ECS650/ECS789 - Database Systems,
  * ECS401U - Procedural Programming (Java).

Ph.D. Research Committee Representative Jun. 2016 - Mar. 2018

- Responsible for the liaison between research groups and the public engagement of our group (e.g. outreach initiatives such as Ph.D. induction and research showcases).

Liguo Steel Group (HK) Limited Hong Kong

Logistic Manager Trainee Jun. 2014 - Apr. 2015

- Worked on commodity (e.g. iron ore, steel) shipping, trading and financing.

Cigna & CMC Life Insurance CO., LTD Shenzhen, China

IT Intern in the Project Management Team Jul. 2012 - Sep. 2012

- Assisted in the software development process and was responsible for tracking issues encountered in user acceptance testing.

JOURNAL PAPER

[1] **Haoyuan Zhang** and D. William R. Marsh, 2018. Generic Bayesian Network Models for Making Maintenance Decisions from Available Data and Expert Knowledge. ***Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability***, 232(5):505-523.

[2] Hua Bai and **Haoyuan Zhang**, 2017. CPN Based Modelling of Tourism Demand Forecasting. ***International Journal of Business and Management***, 12(1): 28-35.

CONFERENCE
PAPER

[3] **Haoyuan Zhang** and D. William R. Marsh, 2018. Towards A Model-Based Asset Deterioration Framework Represented by Probabilistic Relational Models. In *European Safety and Reliability Conference 2018 (**ESREL 2018**)*, pages 671-679 (oral presentation).

[4] **Haoyuan Zhang** and D. William R. Marsh, 2016. Bayesian Network Models for Making Maintenance Decisions from Data and Expert Judgment. In *European Safety and Reliability Conference 2016 (**ESREL 2016**)*, pages 1056-1063 (oral presentation).

[5] **Haoyuan Zhang** and Hua Bai, 2016. Simulation of Tourism Supply Chain Collaborative Demand Forecast. In *International Conference on Applied Social Science Research (**ICASSR 2015**)*, pages 659-662.

UNDER REVIEW

[6] **Haoyuan Zhang** and D. William R. Marsh. Learning from Uncertain Data, Knowledge and Similar Groups: Individualised Multi-State Deterioration Prediction for Infrastructure Asset. Submitted to ***Information Sciences***.

[7] **Haoyuan Zhang** and D. William R. Marsh. Managing Infrastructure Asset: Bayesian Networks for Inspection and Maintenance Decisions Reasoning and Planning. Submitted to ***Reliability Engineering & System Safety***.

[8] **Haoyuan Zhang**, Kaijian Li, Tak Nam Wong, Luping Zhang and Asheem Shrestha. A Colored Petri Net Approach to Aid Integrate Process Planning and Scheduling Optimized by Hybrid Genetic Algorithm and Simulated Annealing. Submitted to ***Expert Systems with Application***.

SKILLS

Programming Language: Java, R, Python, Scala, HTML

Database: PostgreSQL, MySQL

Language: English, Chinese, Cantonese