Qixiang **Fang**

PhD Candidate in Statistics | Natural Language Processing

linkedin.com/in/fqixiang github.com/fqixiang
+31 684 666 253 q.fang@uu.nl
Padualaan 14, 3584 CH Utrecht, the Netherlands
Born in 1992 in Guangdong, China

With an interdisciplinary background in statistics, research methodology, human sciences (e.g. behavioural and medical sciences), software engineering and machine learning (ML), I am particularly interested in the intersection of statistics and ML. Exemplar topics are causal inference with ML models, explanatory ML and high-dimensional measurement models with ML techniques. Furthermore, I am interested in the implications of such research for the human sciences and how they can be translated into open-source software to benefit research.

Skills

R Skills	**Data manipulation** (e.g. dplyr, pipeline, lubridate), **data visualisation** (e.g. ggplot2, plotly, tmap, ImageMagick), **parallel computing** (e.g. foreach), **statistical modelling** (e.g. glm, survival, discsurv, lme4, survey) and **machine learning** (e.g. caret, glmnet, keras)
Python Skills	**Data manipulation** (e.g. Pandas, Numpy), **machine learning** (e.g. Scikit-learn, Keras, Gensim), **web scraping and automation** (e.g. Scrapy, Selenium)
Stats/ML Knowledge	GLM, multilevel, SEM, survival analysis and their Bayesian variants, **statistical learning** (e.g. regularisation, decision trees, ensemble methods, SVM), **deep learning** (CNN, RNN, Han), **language models** (e.g. Word2Vec, BERT), experimental and probabilistic survey designs
Other	JavaScript, SQL, Mplus, HLM, LaTeX, Git, SPSS

Education

2017 - 2019 MSc in Statistics (Cum Laude) at Utrecht University, The Netherlands
2013 - 2016 Joint BA in Psychology & Social Sciences (Cum Laude) at Jacobs University Bremen, Germany

EXPERIENCES

Now – June 2020 | PhD Candidate in Statistics, ML & NLP, Utrecht University, The Netherlands
- Project : Using ML & NLP techniques to tackle validity issues with high-dimensional measurements
- Supervisors : Dr. Daniel Oberski & Dr. Dong Nguyen

Machine Learning | NLP | SEM | Causal Models | Python

May 2020 – September 2019 | PhD Candidate in Software Engineering, Utrecht University, The Netherlands
- Project : Creating and evaluating tools that leverage large volumes of data generated by online educational tools, with the goal to improve learning and teaching.
- Supervisors : Dr. Sergey Sosnovsky & Dr. Johan Jeuring
- Training : Completed a master's level course on machine learning and deep learning.
- Roles : Reviewer for HRI 2020; lecturer of two statistics courses; first supervisor of a bachelor computer science thesis; statistical consultant for colleagues.
- Resignation : Voluntary resignation from the project after 9 months.

Learning Analytics | Software Engineering | Educational Science | R

July 2019 – September 2018 | Research Intern, Statistics Netherlands, The Netherlands
- Project : Understanding and predicting daily web survey response rates using interpretable machine learning models, with predictors derived from weather and Google Trends records.
- Supervisors : Joep Burger, Ralph Meijers & Kees van Berkel
- Output : Best thesis award; manuscript under review at *Survey Research Methods*.

Official Statistics | R | Regularisation | Survival Analysis | Response Modelling

May 2019 – January 2018 | Research Assistant, Utrecht University, The Netherlands
- Data collection, pre-processing, documentation and visualisation for *ASReview*;
- Wrote online statistical tutorials (e.g. *discrete-time survival analysis*, *multilevel GLM*);
- Developed 2D and 3D graphics to facilitate selection of measurement invariance model priors;
- Used latent class analysis to investigate the influence of parenting styles on children's physical and psychological development.

R | Data Visualisation | Literature Research | Latent Variable Modelling

June 2017 – October 2016 | **Data Analyst, WINIT GmbH, Germany**

- Automated update and retrieval of sales data using Python and JavaScript.
- Advised on warehouse storage efficiency and in-house logistics efficiency.
- Data management using Excel, Google Spreadsheets, and SQLite.

Python | Web Scraping | Web Automation | Database Management

August 2016 – June 2015 | **Research Assistant, Jacobs University Bremen, Germany**

- Supervisors : Dr. Margrit Schreier & Dr. Katja Hanke
- Project : The effect of meditation on the well-being of highly sensitive persons.
- Tasks : study design, data collection and data analysis.

Mplus | SPSS | Mediation | Resampling Techniques

Projects & Publications

Project : Classifying Wikipedia Vital Articles Using Deep learning 2019 - 2020

github.com/https://github.com/fqixiang/PatternRecognition Paper

Implemented a CNN in Keras as part of a course group project where we used several neural network architectures to predict the main topic of Wikipedia articles; wrote a substantial part of the project report.

R | Python | Deep Learning | Keras | CNN | LSTM | HAN

Project : ASReview - Active learning for Systematic Reviews 2018 - 2019

github.com/asreview/asreview Website

Collected, pre-processed and documented both open-source and private data sets for the ASReview project; visualised word embeddings in 3D space; tested software.
Output : A co-authored manuscript under review at *Nature Machine Learning*.

Python | Data Collection | Data Pre-processing | Embedding Visualisation

Manuscript : Modelling Web Survey Response Rates Using Time, Weather and Google Trends Data 2018 - NOW

Authors : **Fang**, Burger, Meijers, & van Berkel
Content : We used interpretable machine learning models to predict daily web survey response rates over time, using contextual time-varying predictors constructed from weather and Google Trends data. Our models achieved up to 15% reduction in prediction error compared to the state of the art approach.
Progress : Under review at *Survey Research Methods*; accepted for presentation at BigSurv20.

Interpretable Machine Learning | Adaptive LASSO | Discrete-Time Survival Analysis

Manuscript : Towards Adaptive Social Comparison for Education 2019 - NOW

Authors : Sosnovsky, **Fang**, De Vries, Luehof and Wiegant
Content : We conducted two experiments where we examined the influence of social comparison levels on the learning behaviour and outcomes across different student groups.
Progress : Accepted by ECTEL 2020 as a poster paper.

Personalised Learning | Process Analysis | Digital Traces | Clickstream Analysis

Manuscript : A visualisation of MGCFA, alignment and approximate MI with different priors 2018 - NOW

Authors : Arts, **Fang**, Van de Schoot, & Meitinger
Content : We propose a novel visualisation method to facilitate selection of measurement invariance models.
Progress : Abstract accepted at *Frontiers in Psychology - Quantitative Psychology and Measurement*

Measurement Invariance | Bayesian | Multigroup CFA

Languages

English ●●●●●
Mandarin ●●●●●
German & Dutch ●●○○○

Characters

- Curiosity
- Conscientiousness
- Attention to details

References

Dr. Daniel Oberski
Associate Professor, Utrecht University
@ d.l.oberski@uu.nl
+31 302 539 039

Dr. Dong Nguyen
Assistant Professor, Utrecht University
@ d.p.nguyen@uu.nl
NA