Qixiang FANG

PhD Candidate in Statistics | Machine Learning | Natural Language Processing

in 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, social sciences, software engineering and machine learning (ML), I am particularly interested in topics at the intersection of statistics and ML (e.g. statistical inferences with ML, measurements and causal inferences with textual data). I am also interested in the implications of such research for the social sciences and open-source research (statistical) software.

▤ 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, rpart, keras)
Python Skills	**Data manipulation** (e.g. Pandas, Numpy), **machine/deep learning** (e.g. Scikit-learn, Keras, PyTorch), **NLP** (e.g. NLTK, Gensim, spaCy), **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, ERNIE), probabilistic survey and experimental 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 knowledge from statistics, ML & NLP to improve measurements of and causal inferences with high-dimensional textual data. ➤ Supervisors : Dr. Daniel Oberski & Dr. Dong Nguyen ML NLP SEM Causal Models Python R
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 to join a new PhD project. 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 and predictors derived from weather and Google Trends records. ➤ Supervisors : Dr. Joep Burger, Dr. Ralph Meijers & Dr. Kees van Berkel ➤ Output : Best thesis award ; manuscript under review at *Survey Research Methods*. Official Statistics R Regularisation Survival Analysis Response Modelling Interpretable ML
May 2019 January 2018	**Research Assistant, UTRECHT UNIVERSITY, The Netherlands** ➤ Data collection, pre-processing, documentation and visualisation for *ASReview*. ➤ Developed online statistical tutorials (e.g. *discrete-time survival analysis*, *multilevel GLM*). ➤ Created 2D and 3D graphics to facilitate selection of priors for measurement invariance model. ➤ Used latent class analysis to investigate the influence of parenting styles on children's physical and psychological development for a research team. Data Visualisation Literature Research Latent Variable Modelling Language Models R Python

June 2017 October 2016	**Data Analyst, WINIT GMBH, Germany** › Automated update and retrieval of sales data using Python and JavaScript. › Advised on warehouse storage and in-house logistics solutions based on data-driven insights. › Data management using Excel, Google Spreadsheets, and SQLite. Python │ JavaScript │ 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, data analysis and report writing. 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 multiple neural network architectures (CNN, LSTM and HAN) to predict the main topic of Wikipedia articles; wrote a substantial part of the project report (grade : 9/10).

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; prepared word embeddings and visualised them in 3D space; tested software; supervised a Bachelor thesis about this project.
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 only non-personal, privacy-friendly contextual data. For the predictors, we used time indicators (like day of a week, public holidays), weather records and societal trends indicators (e.g. disease outbreaks, societal concern about privacy issues, travel trends and terrorism salience) which we carefully constructed from Google Trends data. Our models achieved up to 15% reduction in prediction error compared to the state of the art approach, when tested on an independent data set.
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 - 2020
Authors : Sosnovsky, **Fang**, De Vries, Luehof and Wiegant
Content : We conducted two experiments where we examined the influence of social comparison levels on students' learning behaviour and outcomes across different student subgroups. Based on the results, we advised on personalised education.
Progress : Accepted by ECTEL 2020 as a short 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