

Zhaohao Zeng

Tel: (+81) 070-4015-6227 Address: 05-24, Bldg 63, Waseda Univ, Tokyo
E-mail: zhaohao.zeng93@gmail.com GitHub: www.github.com/matthew-z

Experience

Oct 2015 - Present	Silicon Studio Corporation	Tokyo, Japan
	Big Data Engineer	
	Part time(onsite, 28 hours per week). I am participating in Game Data Science research, and integrating Data Science and Machine Learning R/Python scripts into the Spark pipeline, and implementing related API and visualisation tools.	

Education

1, / Sep 2015 - Present	Waseda University	Tokyo, Japan
	Master of Engineering in Computer Science	
	Supervisor: Prof. Tetsuya Sakai	
Sep 2013 - Jun 2015	The University of Liverpool	Liverpool, UK
	Bachelor of Engineering in Electronics	
Sep 2011 - Jun 2013	Xi'an Jiaotong - Liverpool University	Suzhou, China
	Bachelor of Engineering in Electronics Science and Technology	
	Transferred to Liverpool after completing Year-0 & Year-1.	

Publication

Sakai, T., **Zeng, Z.** and Luo, C.:Evaluating Helpdesk Dialogues: Initial Considerations from An Information Access Perspective. IPSJ - Special Interest Group of Natural Language (IPSJ-SIGNL 228) , Osaka, Japan, Sep 2016.

Honours

2015	Monbukagakusho Honours Scholarship
	The Ministry of Education(MEXT), Japanese Government.
2016	Partial Tuition-waiver Scholarship
	Waseda University
2013	Partial Tuition-waiver Scholarship
	The University of Liverpool

Selected Projects

- 2016 **Evaluation Measure of Helpdesk System** Data Crawling & Dialogue Evaluation
Built a helpdesk dialogue test collection by crawling dialogues on Weibo with Python. Built an annotating platform to annotate the dialogues with subjective labels and nugget labels for evaluation via crowdsourcing. Designed a suite of language-independent evaluation measures for quantifying the success of each dialogue.
- 2016 **Google Person Finder** Web Development & Open Source Contribution
Contributed to a Google-owned open-source project called Person Finder, which is based on Python, Django and AppEngine. Improved Person Finder's searching, indexing, and query pre-processing modules to help it handle Chinese and Japanese better.
- 2015 **Cross-Device Targeting in Big Data** Machine Learning & Big Data
Proposed a framework using unsupervised learning algorithms to target user across multiple digital devices, with Apache Spark and Scala. Helped online advertising publisher better track user and recommend products. Utilised a big dataset which was provided by our partner Wider Planet, Inc.
- 2016 **Kaggle: Airbnb New User Bookings** Machine Learning
Attended a data science contest on Kaggle to predict Airbnb new user's first booking destination with R. Handled the incomplete data provided by Airbnb into dummy encoding in this research, and applied a solution based on Gradient Boosting. Received a nearly 0.8 nDCG@5 score with the test data.
- 2015 **Network Analysis of Contributors on Github** Machine Learning & Visualisation
Analysed and visualised the relationships between different programming languages and popular repositories on Github. Conducted the research with Google BigQuery and Apache Spark to handle the large dataset.
- 2014 **Algorithm Automatic Timetable Scheduling Algorithm** NP problem
Developed a genetic algorithm based solution to solve an NP problem: university timetable scheduling problem with C++, Python and Objective-C. Applied parameters tuning and closed pattern mining algorithms to accelerate the processing.

MOOC Certificates

University of Washington	Practical Predictive Analytics
University of Washington	Data Manipulation at Scale
Stanford University	Machine Learning
Stanford University	Algorithms: Design and Analysis
John Hopkins University	The Data Scientists Toolbox
The University of Tokyo	From the Big Bang to Dark Energy
University of Pennsylvania	Probability
University of Illinois at Urbana-Champaign	Pattern Discovery in Data Mining
University of Illinois at Urbana-Champaign	Text Retrieval and Search Engines
University of Illinois at Urbana-Champaign	Text Mining and Analytics