Dr. Boxun Zhang

Sr. Data Scientist at SoundCloud
Ph.D. in Computer Science
Based in Berlin, Germany

+4915229402625
zhangboxun@gmail.com
zhangboxun.github.io
Quora

PROFILE

I recently started working at SoundCloud as a senior data scientist. Previously, I worked as a data scientists at Spotify on various topics: machine learning, user behavior measurement, experimentation, and research collaboration. I obtained my PhD in Computer Science from Delft University of Technology, the Netherlands, where I studied user behavior in Peer-to-Peer file sharing systems.

PROFESSIONAL EXPERIENCE

Senior Data Scientist   **May 2017 – present**

SoundCloud, Germany

Research and develop models, tools, and datasets that are used by other data scientists.

Data Scientist   **Mar. 2013 – Feb. 2017**

Spotify, Sweden

Worked in various areas, including:
- Modeled user retention with machine learning techniques
- Measured and modeled user experience frictions in Spotify's mobile app
- Develope machine learning approach that reveals the cause of effect in A/B testing
- Developed a machine learning model that can predict A/B testing result
- Designed and developed the first user session dataset that was widely used within Spotify
- Built intra-session attribution model that associates streaming behavior to other types of user behaviors
- Built model that predicted the time it takes Spotify's Hadoop cluster (2,000+ nodes) to run out of storage capacity.
- Supported multiple product teams by designing and analyzing A/B tests

(Co)supervised PhD and MSc students for research projects:
- S. Magureanu: *Online Learning of Optimally Diverse Rankings* (PhD internship)
- F. Michelinakis: *Passive Mobile Bandwidth Classification Using Short Lived TCP Connections* (PhD internship)
- G. Mörtberg: *Detection of Streaming Fraud* (MSc internship)
- M. Håstad: *Evaluating Different Sessionization Methods in Music Streaming Context* (MSc thesis)
- E. Aalto: *Learning Playlist Representations for Automatic Playlist Generation* (MSc thesis)
- T. Vasiloudis: *Extending Recommendation Algorithms by Modeling User Context* (MSc thesis)

Research Intern   **Jan. 2012 – May 2012**

Spotify, Stockholm, Sweden

Investigated various aspects of user behavior in Spotify, such as arrival patterns of user sessions and usage patterns of users with multiple devices.

Software Engineer   **Feb. 2006 – Apr. 2008**

International Air Transport Association, Beijing, China

Designed and developed web services for internal systems.

EDUCATION

Ph.D. in Computer Science   **May 2008 – Jan. 2013**

Delft University of Technology, Delft, the Netherlands

- Thesis: Understanding Operation and User Behavior in Peer-to-Peer Systems
- Promotor: Prof.dr.ir. H.J. Sips
- Supervisor: Dr.ir. J.A. Pouwelse

I studied the system dynamics and user behaviors in Peer-to-Peer (P2P) systems like BitTorrent, through empirical analysis and statistical modeling. I was also one of the main contributors of P2P Trace Archive — one of the largest public data archives for P2P traces.

M.Sc. in Computer Science   Aug. 2004 – Feb. 2016

Royal Institute of Technology, Stockholm, Sweden

- Thesis: Implementation of Agora System (multi-agent systems)
- Supervisor: Prof. Mihhail Matskin

B.Sc. in Computer Science   Sept. 2000 – Jun. 2004

Royal Institute of Technology, Stockholm, Sweden

- Thesis: Device Driver Development for Embedded Linux
- Supervisor: Dr. Zhongmei Ma

TALKS
Machine Learning the Product. Nordic Data Science Summit, 2016

Introduction to Locality Sensitive Hashing. Guest lecture for Search Engines and Information Retrieval Systems, KTH, 2016

Data Science at Spotify with Boxun Zhang. Software Engineering Daily, 2015

Music Recommendation in Spotify. Online Social Networks: Data Management and Privacy/Security workshop, 2015

PUBLICATIONS
Stefan Magureanu, Alexandre Proutiere, Marcus, Isaksson, **Boxun Zhang**. *Online Learning of Optimally Diverse Rankings.* ACM SIGMETRICS, 2018 (accepted).

Foivos Michelinakis, Gunnar Kreitz, Riccardo Petrocco, **Boxun Zhang**, and Joerg Widmer. *Passive Mobile Bandwidth Classification Using Short Lived TCP Connections.* IFIP WMNC, 2015.

Boxun Zhang, Gunnar Kreitz, Marcus Isaksson, Javier Ubillos, Guido Urdenata, Johan Pouwelse, and Dick Epema. *Understanding User Behavior in Spotify.* IEEE INFOCOM, 2013.

Boxun Zhang, Alexandru Iosup, Johan Pouwelse, and Dick Epema. *Characterizing Flashcrowds in BitTorrent.* ICT.OPEN, 2012. *Best paper award of ASCI track*

Boxun Zhang, Alexandru Iosup, Johan Pouwelse, and Dick Epema. *Identifying, Analyzing, and Modeling Flashcrowds in BitTorrent.* IEEE P2P Computing, 2011.

Boxun Zhang, Alexandru Iosup, Johan Pouwelse, and Dick Epema. *The Peer-to-Peer Trace Archive: Design and Comparative Trace Analysis.* ACM CoNEXT Student Workshop, 2010.

Boxun Zhang, Alexandru Iosup, Johan Pouwelse, and Dick Epema. *Towards a Peer-to-Peer Trace Archive.* ASCI, 2010.

Boxun Zhang, Alexandru Iosup, Johan Pouwelse, Dick Epema, and Henk Sips. *Sampling Bias in BitTorrent Measurements.* Euro-Par, 2010.

Boxun Zhang, Alexandru Iosup, Johan Pouwelse, and Pawel Garbacki. *A Unified Format for Traces of Peer-to-Peer Systems.* LSAP, 2009.

David Hales, Rameez Rahman, **Boxun Zhang**, Michel Meulpolder and Johan Pouwelse. *BitTorrent or BitCrunch: Evidence of a credit squeeze in BitTorrent?.* COPS Workshop, 2009.

Xuetao Niu, **Boxun Zhang**, and Mihhail Matskin. *An Architecture and System for Support of Cooperation in Multi-Agent Software Development.* SERP, 2007.

SKILLS
Programming: Python, R, Scala, Java

Big data: MapReduce, Spark, BigQuery, Hive, SQL

Data visualization: ggplot2, Seaborn, Matplotlib, D3.js, Graphviz, Gnuplot

LANGUAGES
English – fluent/proficient in speaking, reading, and writing

Chinese – native speaker