DAYI LIN, Ph.D.

Software Engineering Researcher · Data Scientist

CONTACT

+1-(613)-770-8659   dayi.lin@cs.queensu.ca   lindayi.me

linkedin.com/in/dayilin   RG researchgate.net/profile/Dayi_Lin

HIGHLIGHTS

➢ Proven effectiveness in both industry (experienced industrial data scientist) and academia (6 flagship journal publications)

➢ Prestigious media coverage including Kotaku, PC Gamers, and national newspapers

➢ Research interests: software engineering, mining software repositories, game engineering, defect prediction, software engineering for machine learning systems

EDUCATION

Ph.D. – Queen's University, Canada

Sept. 2015 – Jan. 2019 | Computer Science | Ultra-Large Software System Specialization

➢ Supervisor: Prof. Ahmed E. Hassan (Software Analysis and Intelligence Lab)
➢ Thesis Topic: *How Can Game Developers Leverage Data from Online Distribution Platforms? A Case Study of the Steam Platform*

B.Eng. – Xi'an University of Posts and Telecommunications, China

Sept. 2011 – Jun. 2015 | Computer Science and Technology | 98% Percentile

PUBLICATIONS

(PEER-REVIEWED)

Journal Publications

An Empirical Study of the Characteristics of Popular Minecraft Mods

➢ D Lee, GK Rajbahadur, **D Lin**, M Sayagh, CP Bezemer, AE Hassan
➢ *Empirical Software Engineering (EMSE)*, 2019 (Impact Factor 4.457)

Building the Perfect Game – An Empirical Study of Game Modifications

➢ D Lee, **D Lin**, CP Bezemer, AE Hassan
➢ *Empirical Software Engineering (EMSE)*, 2019 (Impact Factor 4.457)

Identifying Gameplay Videos that Exhibit Bugs in Computer Games

➢ **D Lin**, CP Bezemer, AE Hassan
➢ *Empirical Software Engineering (EMSE)*, 2019 (Impact Factor 4.457)

An Empirical Study of Game Reviews on the Steam Platform

➢ **D Lin**, CP Bezemer, Y Zou, AE Hassan
➢ *Empirical Software Engineering (EMSE)*, 2018 (Impact Factor 4.457)

An Empirical Study of Early Access Games on the Steam Platform

➢ **D Lin**, CP Bezemer, AE Hassan
➢ *Empirical Software Engineering (EMSE)*, 2017 (Impact Factor 4.457)

Studying the Urgent Updates of Popular Games on the Steam Platform

➢ **D Lin**, CP Bezemer, AE Hassan
➢ *Empirical Software Engineering (EMSE)*, 2017 (Impact Factor 4.457)

Conference Publications

Chinese Word Segmentation based on Conditional Random Fields with Character Clustering

➢ L Du, X Li, C Liu, R Liu, X Fan, J Yang, **D Lin**, M Wei
➢ *The 20th International Conference on Asian Language Processing (IALP)*, 2016

Chinese Term Extraction from Web Pages based on Expected Point-wise Mutual Information

➢ L Du, X Li, **D Lin**
➢ *The 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, 2016

Thesis

How Can Game Developers Leverage Data from Online Distribution Platforms? A Case Study of the Steam Platform

➢ *Ph.D. Thesis*, 2019

ACADEMIC EXPERIENCE

Research Assistant – Software Analysis and Intelligence Lab, Queen's University

Sept. 2015 – Jan. 2019 | Mining Software Repositories, Game Engineering

➢ Conducted research on the modern software development lifecycle of a unique software genre – PC games

➢ Combined both qualitative methods (e.g., Grounded Theory) and quantitative methods (e.g., Natural Language Processing and Machine Learning) to uncover patterns and insights in game store data. Proposed practical suggestions that help game developers produce higher quality games and improve user satisfaction

➢ Built an intelligent tool to automatically identify game bugs from gameplay videos, with a median average precision at 10 of 0.91

➢ Improved the interpretation and goodness-of-fit by 100% for just-in-time cross-project defect models using context-aware mixed-effect modelling

Research Assistant – Text Mining Group, XUPT

Sept. 2012 – Jul. 2015 | Natural Language Processing, Network Analysis

➢ Designed and developed a visualization and query engine for knowledge graphs

> Proposed a novel approach to rank entities in an academic network by constructing and mining a large-scale heterogeneous information network

---

PROFESSIONAL EXPERIENCE

Data Scientist – Prodigy Game

Feb. 2019 – Present | Statistical Modelling, Machine Learning, Experiment Design

> Conduct research on player behavior modelling; model complex in-game user behavior at both user and session levels, with data from 70+ million users who generate 300+ million events per day

> Conduct research on player segmentation using both quantitative and qualitative approaches (unsupervised machine learning + user survey); profile players based on their in-game behavior to empower personalized gaming experiences

> Analyze key engagement features and their impact on key metrics (e.g., user retention); provide in-depth data-driven insights that power product iterations

> Collaborate with product and user research teams to design, conduct and analyze in-game experiments for new releases

Data Researcher – Ford Motor Company

Jan. 2017 – Apr. 2017 | Algorithm Design, Statistical Modelling

> Designed Markov Chain-based algorithms, in combination with binary classifiers to identify orphaned or wrongly recognized voice commands, and mistakes in user behaviors (e.g., mis-clicks)

> Analyzed user behavioral data from infotainment systems in Ford vehicles to provide feedback to design and development teams, improving the user-perceived quality of the system

> Instrumented the source code of a legacy large-scale distributed system and used log mining techniques to collect and stream necessary data for online analysis

Data Researcher – Blackberry

Jan. 2017 – Apr. 2017 | Computational Data Analysis, Statistical Modelling

> Member of the Location and Telemetry Technologies team, which is responsible for both collecting and processing analytics data from all BlackBerry platforms in order to track feature usage and improve future products

Intermediate Research & Development Engineer – Tmall, Alibaba Group

Jun. 2014 – Sept. 2014 | Back-end Web Development, Java

> Developed the backend and internal tools for Member Entertainment System; developed the backend of the membership module of Tmall (one of the largest ecommerce websites in China), which supports millions of pageviews per day

TEACHING EXPERIENCE	**Academic Assistant – Smith School of Business, Queen's University**

Aug. 2019 – Present | MMA-865: Big Data Analysis, MMA-869: Machine Learning & A.I.

- ➢ Prepare course materials for the Big Data Analysis and Machine Learning & A.I. courses
- ➢ Give lectures about using Spark and Spark MLlib to process and model data at scale

Head Teaching Assistant – School of Computing, Queen's University

Sept. 2016 – Dec. 2018 | CISC-326: Software / Game Architecture

- ➢ Prepared course materials for the Software / Game Architecture course
- ➢ Gave lectures about software requirement engineering and modern code analysis tools (*Understand*)
- ➢ Led four other teaching assistants in organizing the 200-student class

---

MENTORED STUDENT

Daniel Lee (2017 – 2019, Research-based M.Sc. student)
- ➢ Thesis Topic: *Studying the Distribution Platforms for Game Mods to Create Active Modding Communities*

PRESENTATIONS

An Empirical Study of Game Reviews on the Steam Platform
- ➢ *Invited Talk, Workshop on Video Game Criticism, Université de Montréal*, Nov. 2018

An Empirical Study of Early Access Games on the Steam Platform
- ➢ *Journal First Presentation, 40th International Conference on Software Engineering (ICSE 2018)*, May 2018

Studying the Urgent Updates of Popular Games on the Steam Platform
- ➢ *Journal First Presentation, 33rd International Conference on Software Maintenance and Evolution (ICSME 2017)*, Sept. 2017

Research on Games: Studying the Steam Platform
- ➢ *Guest Lecture, CISC-880: Mining Software Repositories, School of Computing, Queen's University*, Oct. 2016

Research on Games: Studying the Steam Platform
- ➢ *Guest Lecture, CISC-326: Game Architecture, School of Computing, Queen's University*, Oct. 2016

SELECTED MEDIA COVERAGE

- ➢ **Kotaku:** Big study of 10 million Steam reviews is absolutely fascinating (http://bit.ly/2u1n9La)

- ➢ **Kotaku:** Research Paper on Steam Early Access Reveals 5 Lessons For Developers (http://bit.ly/2OaA6t7)
- ➢ **PC Gamer:** Steam study finds users leave negative reviews more quickly than positive ones (http://bit.ly/30xfwbG)
- ➢ **Rock Paper Shotgun:** Study suggests Steam reviewers are bothered more by bad game design than bugs (http://bit.ly/389Pde1)
- ➢ **Gamasutra:** Steam reviews study suggests 'bad design', not bugs, irks players most (https://ubm.io/2Nwbwmg)
- ➢ **Der Standard:** Studie zu Steam-Reviews: User sudern schneller, ärgern sich kaum über bugs (http://bit.ly/35SYHZz)

TECHNICAL SKILLS

Domains

Software Engineering   Mining Software Repositories   Data Science   Machine Learning

Software Engineering for Machine Learning   Machine Learning for Software Engineering

Research Methods

Quantitative & Qualitative Analysis   Grounded Theory   Experimental Research Design

Programming Languages

Python   R   Java   C/C++   PHP   SQL   LaTeX   HTML/CSS

Frameworks

Spark   MLlib   GraphX   Kafka   Hadoop   Numpy/Scipy   Scikit-Learn

---

HONORS & AWARDS

2019	**Invited Participant**, Shonan Meeting No. 156: Software Engineering for Machine Learning Systems, NII Japan
2018	**Invited Participant**, Workshop on Video Game Criticism, Université de Montréal
2015-2019	**Recipient**, Queen's Graduate Award, Queen's University
2017-2018	**Recipient**, Conference Travel Award, Queen's University
2018	**Winner** and **Avanade Sponsored Prize Recipient**, QHacks
2016	**Winner**, Limestone City Hacks
2015	**Outstanding Graduates**, Xi'an University of Posts and Telecommunications

---

REFERENCES

Dr. Ahmed E. Hassan

Professor, Queen's University

Email: ahmed@cs.queensu.ca

Website: http://sail.cs.queensu.ca

Dr. Cor-Paul Bezemer

Assistant Professor, University of Alberta

Email: bezemer@ualberta.ca

Website: https://www.ece.ualberta.ca/~bezemer/