Li-An Yang

l3yang@ucsd.edu (858) 344-2107
linkedin.com/in/li-an-yang
li-an.me

Skills

Area of Interest	Recommender System, Data Mining, NLP, Predictive Analytics
Programming Languages	Python, Java, SQL, Shell Script, C++
Miscellaneous	BPR, Collaborative Filtering, XGBoost, BoW, Docker, IBM Watson, scikit-learn

Education

University of California, San Diego	M.S. in Computer Science	Sep. 2018-Exp. Mar. 2020
National Chiao Tung University	B.S. in Computer Science	Sep. 2012-Jun. 2016
Université de Technologie de Compiègne	Exchange Student	Feb. 2016-Jun. 2016

Selected Term Projects (li-an.me#works)

Amazon Purchase Prediction Challenge on Kaggle

- Ranked 7th among 800 people with an accuracy of 72.96%
- Formulated the problem as a ranking objective function to maximize pairwise preference prediction probability
- Adopted collaborative filtering techniques by exploiting category and popularity similarities between items

A Preliminary Study of Automatic Playlist Continuation and Add-to-playlist Prediction on Spotify

- Designed two variations of feature vector for track-level and playlist-level clustering respectively
- Recommended on average 9% of total tracks to cover 86% of the ground truth dataset
- Suggested playlists for users to add a liked track to with logistic regression models and XGBoost

Automatically Proving Mathematical Theorems with Evolutionary Algorithms and Proof Assistants

- Research results published at the IEEE Congress on Evolutionary Computation
- The first to generate formal proofs automatically by exploiting proof assistant Coq with evolutionary algorithms
- Proved ten theorems in different branches of mathematics automatically: Arithmetic, Logic, & Parity

A Voice-controlled Streaming Jukebox based on IBM Bluemix Cloud Service

- Established music streaming services on Raspberry Pi featuring personal music recommendation
- Deployed IBM Watson APIs to carry out on-demand speech-to-text features and social networking services

Right Whale Recognition Competition on Kaggle

- Adopted SIFT and bag-of-words model to extract distinctive feature of the whale face in a team of three
- Improved the evaluated score in log-loss by thirty percent in limited time

Work Experience

Software Engineer Intern, Verizon Media — Jun.2019-present

Research Assistant, Institute of Information Science, Academia Sinica — Mar. 2017-May 2018

- **User Intention Understanding on Taiwan Open Platform for Educational Resources**
  - Revealed search keyword trends on the website with millions of entries of user log data
  - Developed a keyword generation pipeline for personalized resource recommendation in primary education
- **SQUAT - a Sequencing QUality Assessment Tool in Bioinformatics (**github.com/luke831215/squat**)**
  - Tools released on Github and Docker Hub and results submitted to an open-access, peer-reviewed journal
  - Provided read mapping analytics and visualized the assessment results in a portable HTML report
  - Devised metrics for researchers to identify poor-quality data upon terabytes of data gathering

R&D Intern, Email Reputation Services Division, Trend Micro Inc. — Jun. 2015-Aug. 2015

- **Pattern Recognition**
  - Simulated patterns of malicious mail attacks with log-based data hashing and clustering
  - Reduced volume for processing by designing a white-listing mechanism to filter out known legitimate entries

Publications (li-an.me#pubs)

1. **L.-A. Yang**, Y.-J. Chang, S.-H. Chen, C.-Y. Lin, & J.-M. Ho. “SQUAT: A Sequencing Quality Assessment Tool for Data Quality Assessments of Genome Assemblies.” Journal, admitted to BMC Genomics (2019)
2. **L.-A. Yang**, W.-C. Chung, Y.-J. Chang, S.-H. Chen, C.-Y. Lin, & J.-M. Ho. “The Spiral Assembler: An Iterative Process of NGS De Novo Genome Assembly with Machine-Learning for Subset Selection on Quality-Score and K-Mer Landscape.” Technical Report, submitted to Institute of Information Science, Academia Sinica. doi: TR-IIS-18-001.
3. **L.-A. Yang**, J.-P. Liu, C.-H. Chen, & Y.-P. Chen. “Automatically Proving Mathematical Theorems with Evolutionary Algorithms and Proof Assistants.” In Proceedings of 2016 IEEE Congress on Evolutionary Computation (CEC 2016). (pp. 4421–4428). doi: 10.1109/CEC.2016.7744352. (EI). (github.com/nclab/ea.prover)