Zheng Xie

(+86)156-0519-5337 | xie.zheng777@gmail.com | lamda.nju.edu.cn/xiez | Tedxz | zhengxie

Currently, I am a Ph.D. student in Nanjing University, LAMDA Group. My research interests lie in Machine learning and some related field, including semi-supervised learning and imbalanced learning. I have several research papers published on top conferences and journals. I received my B.Eng degree in Computer Science and Technology in June 2016 from Xi'an Jiaotong University. During the undergraduate period, I devoted myself into algorithm competitions like ACM-ICPC, and interesting opensource projects.

Education

Nanjing University, LAMDA Group — *Nanjing, China*

Ph.D. student in Computer Science, Working on Machine Learning — *Sept. 2016 - Present*

- Supervised by Prof. Zhi-Hua Zhou and Prof. Ming Li.
- Research interests lie in machine learning, especially in semi-supervised learning and imbalanced learning.

Xi'an Jiaotong University — *Xi'an, China*

B.Eng. in Computer Science — *Sept. 2012 - July 2016*

- Served as a co-coach for competitive programming.
- My undergraduate thesis was named outstanding thesis.

Experience

Research Intern at Baidu Research — *Beijing, China*

Big Data Lab, Baidu Research — *July 2015 - Oct. 2015*

- Mining from spatial-temporal big data.
- User analysis and modelling.
- Human traffic modelling and prediction.

Teaching Assistant — *Nanjing, China*

Nanjing University — *Feb. 2017 - June 2017*

- Basics of Programming, Spring, 2017.
- Introduction to Data Mining, Spring, 2017.

Publications

Cutting the Software Building Efforts in Continuous Integration by Semi-Supervised Online AUC Optimization — *Stockholm, Sweden*

The 27th International Joint Conference on Artificial Intelligence (IJCAI-18) — *2018*

- **Zheng Xie**, Ming Li
- Continuous Build Outcome Prediction is a common approach to cut the effort in Continuous Integration. However, it has four unique challenges that are not properly handled by existing methods, which are 1) streaming data, 2) few build outcome labels, 3) imbalance in build outcomes, and 4) suspiciousness of build outcome event required. In this paper, we propose Semi-supervised Online AUC Optimization that can address the four challenges simultaneously, and outperform the existing methods.

Semi-Supervised AUC Optimization without Guessing Labels of Unlabeled Data — *New Orleans, USA*

The 32nd AAAI Conference on Artificial Intelligence (AAAI-18) — *2018*

- **Zheng Xie**, Ming Li
- Semi-supervised learning is widely used when the labeled data is expensive to collect, and AUC optimization is a standard approach to learn classifiers for ranking. In this paper, we argue that, in semi-supervised AUC optimization, it is unnecessary to guess the possible labels of the unlabeled data or prior probability based on distributional assumptions. We analytically prove the unbiasedness, and propose two semi-supervised AUC optimization methods based on it.

Cost-Sensitive Margin Distribution Optimization for Software Bug Localization — *Tianjin, China*

China Conference on Machine Learning (CCML-17) / Journal of Software — *2017*

- **Zheng Xie**, Ming Li
- Software Bug Localization aims to localize buggy program files from numerous source files, which can reduce the cost of software maintenance. Such a problem suffers from highly imbalanced data. In this paper, we propose a novel cost-sensitive loss function for optimizing margin distribution, with a deep model to extract the semantic information of the source code. Our method improves the localization accuracy by a large margin.

Music Style Analysis among Haydn, Mozart and Beethoven: an Unsupervised Machine Learning Approach *Shanghai, China*

International Computer Music Conference (ICMC-17) *2017*

- Ru Wen, **Zheng Xie**, Kai Chen, Ruoxuan Guo, Kuan Xu, Wenmin Huang, Jiyuan Tian, Jiang Wu
- In this paper, we propose an approach to extract melody based features from sheet music, as well as an unsupervised clustering method for discovering styles from the music. The analysis results of our method conform to the Implication-Realization model, one of the most significant modern theories of melodic expectation, which confirms the validity of our approach.

Honors & Awards

Competition Awards

2015	**Silver Prize**, ACM-ICPC Asia Regional Contest	*Shanghai, China*
2014	**Silver Prize**, ACM-ICPC Asia Regional Contest	*Guangzhou, China*
2014	**First Prize**, China Undergraduate Mathematical Contest in Modeling	*Xi'an, China*
2013	**Gold Prize**, ACM-ICPC China Provincial Contest	*Xi'an, China*
2013	**Gold Prize**, ACM-ICPC China Provincial Contest	*Chengdu, China*
2012	**Bronze Prize**, ACM-ICPC Asia Regional Contest	*Tianjin, China*
2011	**Bronze Medal**, National Olympiad in Informatics	*Changchun, China*

Honors

2018	**AAAI-18 Student Scholarship**, AAAI	*New Orleans, USA*
2017	**First-Class Scholarship**, Nanjing University	*Nanjing, China*
2016	**Outstanding Graduate**, Xi'an Jiaotong University	*Xi'an, China*
2016	**Outstanding Undergraduate Thesis Award**, Xi'an Jiaotong University	*Xi'an, China*

Skills

Machine Learning Machine learning is my main research interest. Currently, I'm working on research involving imbalanced learning and semi-supervised learning. I have several research papers published on academic conferences and journal.

Algorithms Years of experience in algorithm programming contests, since senior school. Won prizes for many times in ACM-ICPC and NOI. Got top 1000 in Google Code Jam, top 500 in Baidu A*, top 0.02% in CCF-CSPRO, etc.

Programming

Python	for machine learning projects and so on,
C++	with many years of experience in competitive programming,
JAVA	for some projects (a tank game, Wechat back-end, etc.),
LaTeX	for technical writing, and maintaining a thesis template for XJTU.

Languages **Chinese** (native), **English**.