KEN TU

Github: HuyTu7 | **LinkedIn:** huytu | **Website:** kentu.us | **Cell:** 919-961-8256 | **Email:** hqtu@ncsu.edu

SUMMARY

- Passionate about the synergy of human and AI to improve software development specifically and socio-technical ecosystems generally.
- Extensive individual & collaborative work in hyper-parameters tuning, NLP/NLU, and recommendation systems.
- Interested in infrastructure development and research-oriented machine learning (or data science) positions.

EDUCATION

NORTH CAROLINA STATE UNIVERSITY (NCSU) *Raleigh, NC*

Ph.D. in Computer Science — Expected: Dec 2021

- Advisor : Dr. Tim Menzies (h-index=57) @ RAISE Lab (Real-world Artificial Intelligence for Software Engineering)

M.S. in Computer Science — May 2019

APPALACHIAN STATE UNIVERSITY (ASU) *Boone, NC*

B.S. in Computational Mathematics, *magna cum laude* - GPA: 3.80 / 4.0 — May 2016

PROFESSIONAL EXPERIENCES

Computer Science Department @ NCSU *Raleigh, NC*

RESEARCH ASSISTANT — August 2017 - Present

- AI4SE: Research & build tools that are human-focused/explainable AI to better software development.
- SE4AI: Conduct qualitative and quantitative studies to understand how SE processes/philosophies can improve AI.
- Current Projects: (1) NSF SI^2 applying empirical SE for computational science projects to improve software quality of non-traditional software development, and (2) building ML models that makes sense to human for SE tasks (e.g. why did this specific code fails) to increase the efficiency of software development.

TEACHING ASSISTANT — August 2016 - Present

- Coordinate with the professor & other teaching assistants as a team to structure the course (SE, Parallel Architecture, AI, Data Structures), design tests, conduct review sessions, facilitate labs, and deliver the lesson effectively.

Amazon.com Services, LLC *Pittsburgh, PA*

APPLIED SCIENTIST INTERN — May 2020 - August 2020

- Multilingual Natural Language Understanding: utilized Tensorflow to explore cross-lingual transferring approaches and expand the current monolingual pipeline to multilingual which boost the performance of low-resource languages and production candidate model up to *15% with just 50K data instances*.
- The work is integrated into production & was documented as a research paper for an internal research conference.

Pinterest Inc. *San Francisco, CA*

MACHINE LEARNING ENGINEER INTERN — May 2019 - August 2019

- Users' interest recommendation: explored and built such prototype from *300k+ users' activities* (via Presto & Hive) to boost users' engagements & serve as features candidate generation for downstream functions (ads, homefeed, etc)
- Found biases within the existing models (PCA, SVD, NMF, & ALS) and designed one that is more *diverse (55%)* and *relevant (33%)* with temporal supervised learning (SVM via Scikit-learn & LSTM via Keras).

YouNet Inc. & Gumi Inc. *Ho Chi Minh City, Vietnam*

R&D DATA SCIENCE INTERN — May 2017 - August 2017

- Devised and enhanced core products (via Python, Spark, and Keras) to analyze social-media contents.
- Facebook User's Age Classifier: investigated unstructured *25k+ Vietnamese Facebook users'* posts with CNN & LSTM to predict users' age. Simple KNN achieved comparable *accuracy=80%* but with *300%+ CPU efficiency*.
- Facial Beauty Rater: is developed based on the SCUT-FBP dataset (500 people images) and OpenFace software to extract facial landmarks as features. CNN outperformed RF (by *20%+ in Pearson coefficient*).

Mathematics & Computer Science Departments @ ASU *Boone, NC*
UNDERGRADUATE RESEARCH ASSISTANT August 2012 - August 2016

- Led and collaborated with professors on *6+ research projects* to prove mathematical theorems (e.g. graph theories and operation research), analyze statistical relationships, and prototype models (via Python, Java, R, and MySQL).

University Housing @ ASU *Boone, NC*
RESIDENT ASSISTANT (RA) & VP of RA COUNCIL January 2013 – August 2015

- Organized educational and social cross-building wide programs that fostered community for *900+ students*.
- Expressed the voice of resident assistants and students on campus through composing policy and legislation proposals from Resident Assistant Council to University Housing Leadership (*13 proposals in 1 year*).

PUBLICATIONS & RESEARCH PROJECTS

The Changing Nature of Computational Science (CSc)

- [Submitted for ICSE conference], as part of NSF SI^2. Through the empirical investigation of *13* current beliefs about SE practices within *59 CSc Github projects*, we can only *endorse 3 beliefs*. This implies (1) what kinds of tools we would propose to better support computational science and (2) research directions for both communities.

Detecting Anomalous Scientific Workflow Data Transfers

- [Submitted for ESA journal], as part of NSF SI^2. An anomaly detector, X-FLASH, identifies faulty TCP signatures in Scientific Workflows (SW). X-FLASH outperformed SOTA up to *40% relatively in recall within 30 evaluations*. This implies (1) a proof-of-concept of tuning SE method for SW and (2) the necessity of tuning for future SW work.

Identifying Self-Admitted Technical Debts with Jitterbug: A Two-step Approach

- [Submitted for TSE journal] Separated SATDs as hard and easy TDs to find *close 100% of easy TDs* while being able to find hard TDs more efficiently (with less human effort) than the prior state of the art methods.

Data Labelling with EMBLEM (and how that Impacts Defect Prediction)

- [Accepted for TSE], as part of NSF SI^2. A novel system with human + AI partnership (incremental SVM active learning) to label buggy commits *8 times faster* and help build defect predictors *78% more accurate*.

Can you Explain that Text, Better? Comprehensible Text Analytics for SE Applications

- [Revising for Journal Submission] Novel combination method of LDA topic modeling and Fast Frugal Tree (depth of 4) to predict the severeness of bug reports in software analytics. Offers comparable performance but is more actionable than the common state-of-the-art text mining methods (TFIDF+SVM and LDADE+SVM).

Is One Hyperparameter Optimizer Enough?

- [Accepted for IEEE SWAN at FSE] Empirical case study for hyperparameter tuning in software defect prediction.

AWARDS AND HONORS

ACM Richard Tapia – Diversity in Computing, 2018-19 | Scholar, IBM and AccessSTEM sponsored (17% acceptance)
ACM Joint ESEC/FSE Keynote, 2018 | Keynote Co-author for Top-tier SE conference
Pi Mu Epsilon Mathematics Honor Society, 2013-Present | Academic Excellence, *top 5%* of the class
Graduate Merits Fellowship, 2015-16 | Notable Mathematics Graduate Student *($10,000+)*, ASU
Who's Who Among Students in American Universities, 2015-16 | National Recognition for Outstanding Student Leader

SERVICE

Research Program
ROSE @ ICSE COMMITTEE MEMBER 2019-20
IEEE EMSE & TSE REVIEWER 2019
ICSE and FSE STUDENT VOLUNTEER 2018

Asian Students in Alliance @ NCSU *Raleigh, NC*
COMMUNITY LIAISON Summer 2017 – Summer 2019

East Coast Asian American Student Union (ECAASU)
DIRECTOR OF ADVOCACY Summer 2016 – Summer 2019