Chuan-Wei (Auberon) Kuo

auberonkuo@gmail.com | linkedin.com/in/auberonkuo | (+886) 922-616-302

R&D Engineer at ITRI (2016.09-2020.02) and currently a Ph.D. candidate (ABD) at NYCU CS Department. Interested in Data Scientist or Machine Learning related opportunities

Place and date of birth: Hsinchu/Taipei (Taiwan), 7th of January 1991

EDUCATION

National Yang Ming Chiao Tung University **Sep 2018 – June 2023**

Ph.D. (ABD) Computer Science and Engineering GPA: 4.09/4.3

National Chung Hsing University **Sep 2014 – June 2016**

M.S. Technology Management GPA: 3.97/4.3

National Kaohsiung Normal University **Sep 2010 – June 2014**

B.S. Applied Mathematics GPA: 3.5/4.0

WORK EXPERIENCE

Software Engineer/Machine Learning Engineer | ITRI **Sep 2016 – Feb 2020**

Computational Intelligence Technology Center

- Programming Language: 80% Python, 20% Java
- Algorithm Development (40%)
- Data Engineer/Analysis (40%)
- Data Analysis (Statistical, Machine Learning, Deep Learning) (20%)
- Env.: Linux, Pycharm, Pytorch, Hadoop (HDFS), Hive, Pyspark, Docker, Jupyter Notebook, MS SQL-Server, VirtualBox, Anaconda3, Eclipse

Marketing Intern | Yahoo! Inc. (Oath Inc.) TW **Sep 2015 – Jan 2016**

- Expanded 158 Tumblr hosters, 1,784 posts and get 27% of active users in the last month of the project
- Won the Best Team in the end of the campaign "Telescope Project"
- Business Model Development

ACCOMPLISHMENTS

PUBLICATIONS:

- ***CNet: A Node Calibrate Network for Citation Graph Classification***
  (Under Review) | IEEE International Conference on Data Mining (ICDM2022)
- ***Learning from Citation: Interpretable Patent Valuation via Conditional Variational Autoencoder***
  (Under Review) | Knowledge and information systems journal (Springer)
- ***Optimized data de-identification using multidimensional k-anonymity***
  Aug 2018, NY, U.S. | IEEE TrustCom/BigDataSE 2018 (Oral)
- ***Investigating the Trend of Cross-Disciplinary Evolution in Modern Technological Development***
  Jul 2016, Cambridge, UK | R&D Management conference 2016 (Oral)
- ***Investigating the Trend of Cross-Disciplinary Evolution in Technology Development***
  Jun 2016, Master Thesis

- ***Exploring Influence of R&D Investment, Import and Export Performances to Patent Value***
  Sep 2015, Portland, U.S. | PICMET 2015 (Oral)
- ***Macroscopic Knowledge Source Analysis for Characterizing Global Industry-University Collaboration***
  Sep 2015, Hsinchu, TW | CSMOT 2015 (Oral)

PATENT (Active):

- ***Data de-identification method, data de-identification apparatus and non-transitory computer readable storage medium executing the same***
  TWI644224B, Publication by Taiwan Intellectual Property Office (TIPO)
  US10699029B2, Publication by U.S. Patent and Trademark Office (USPTO)
  CN109684862B, Publication by National Intellectual Property Administration (CNIPA)

SPECIALTIES

1. **AI Research (Main):** Patent Valuation Prediction by Neural Network (CVAE/GNNs)
2. **AI Research (Touched):** NLP, CV, RL, RNNs
3. **Programming (Main):** Python, SQL
4. **Programming (Touched):** Java, Hive
5. **Development Environment (Main):** Python IDE
6. **Development Environment (Touched):** Java IDE

Experienced development environment and tools

OS: Linux, iOS, Windows

IDE: PyCharm, Eclipse, and Sublime Text 3, Jupyter Notebook

Machine Learning Framework: PyTorch (primary), TensorFlow (touched)

Distributed Ecosystem: Hadoop (HDFS, Hive, Hue), Spark

Container: Docker, Virtual Box, VMware

Database: Hive, MS SQL-Server, MySQL, SQLite, PostgreSQL

Code Management: GitHub, Bitbucket

Web Platform:

- Application (Jinja2 with Python), Web server (Nginx), DB (PostgreSQL), OS (Linux)
- Flask-Redis-Celery

Others: Anaconda3, SSH, Putty, winSCP (FTP transfer), Slack, Trello, Offices

PhD Research

2020-Now

Topic & Status	Experiment	Contribution
Topic: Learning from Citation: Interpretable Patent Valuation via Conditional Variational Autoencoder **Status:** (Under Review) Knowledge and information systems journal (Springer)	**Data:** U.S. Patent and Trademark Office (USPTO) 6,000,000+ patents from year between 1976 and 2018. Also 80,000,000+ citation data. **Source:** https://github.com/PatentsView **Model:** Utilizing existing Conditional Variational Autoencoder (CVAE) model **Baseline:** lightGBM, XGBoost, LR, Ridge, DTree, Gradient Boosting **Metric:** MAPE, MSE, ACC	1. First time use USPTO data on CVAE model 2. The way that utilizes generating model for prediction task (currently generative model focus on graph generation) 3. Taking patent value analysis into a whole new field
Topic: CNet: A Node Calibrate Network for Citation Graph Classification **Status:** (Under Review) 2022 IEEE International Conference on Data Mining (ICDM2022)	**Data:** Cora, Citeseer, Pubmed **Source:** https://github.com/kimiyoung/planetoid **Model:** New model called CNet (Graph Neural Network based) **Baseline:** GraphSAGE, MoNet, GCN, GCNII, GAT, CGNN, GRAND, GPN, GDE **Metric:** ACC, F1-score	1. We propose the idea of implicit and explicit neighbors. They learn the associations between nodes, and the relationships of the graph structure 2. We design the homogeneity neighbors that extract information between nodes and the mapping neighbors to discover the node's citation relationship 3. We propose the CNet model that utilizes implicit and explicit neighbors to update node representations 4. We provide a general node calibrate algorithm that helps improve the efficiency of getting valuable information from citation graphs

PhD Research

2020-Now

Topic: GRNet: Graph Reassemble Network for Patent Valuation Task **Status:** (Working) 2023 IEEE International Conference on Data Engineering (ICDE 2023)	**Data:** Cora, Citeseer, Pubmed, USPTO **Source 1:** https://github.com/kimiyoung/planetoid **Source 2:** https://github.com/PatentsView **Model:** New model called GRNet **Baseline:** GIN, GraphSNN, GraphSAGE, MoNet, GCN, GCNII, GAT, GRAND, GPN **Metric:** MSE, MAPE, ACC, F1-score	1. We propose the GRNet model that further utilizes citation linkage of citation graphs than CNet to update node representations 2. This is first time use USPTO data on GNN prediction task 3. We extract the structural information from patent document
Topic: BERTcard: BERT for Causal Relationship Diseases Finding **Status:** (Working) 2023 International Joint Conference on Artificial Intelligence (IJCAI 2023)	**Data:** Taiwan Health Insurance Information (Not open source) **Model:** New model called BERTcard **Baseline:** RNN, LSTM, MedBERT, BERT (Currently decide, which will increase) **Metric:** -- (Currently decide, which will increase)	1. We propose the BERTcard that utilizing the BERT structure (many embedding results on different layers) to learning the relationship between diseases 2. This is a very early trial of medical data, and first time experiment on Taiwan Health Insurance Information data
Topic: Untitled **Status:** (TBD)	**Data:** Patents and patent graphs from USPTO Patent Full-Text and Image Database. **Source 1:** https://github.com/PatentsView **Source 2:** https://patft.uspto.gov/netahtml/PTO/search-bool.html **Model:** -- **Baseline:** CNN, GAN, ResNet, CVAE (Currently) **Metric:** --	We expect the model can be used for patent infringement prevention via discriminating the similarly patent graph between exist patent graph and new grant patent graph (Similar to flaw detection)

PROJECTS

2016

Project Name	Environment	Specific Work
1. *Big data analysis platform construction and industry consulting services* 2. *Intelligent Business - Consumer behavior and community observation technology*	**OS:** Linux, Windows **Distributed Ecosystem:** Hadoop(HDFS, Hive, Hue) **Programming:** Python, Java **IDE:** PyCharm, Sublime Text 3 **Database:** Hive, MS SQL-Server **Container:** Virtual Box	✧ Understand privacy policies, privacy risk principles, risk formulas, algorithms ✧ Processing the data (up to 1 TB) which store on the HDFS (Hadoop System about 20 nodes) through HIVE ✧ Extension 1. Execute Java API to process data and using sh script execution through Hive (when data is too large) [From 6 different organization] ✧ Patent development (private processing method for time series data)

PROJECTS

2017

Project Name	Environment	Specific Work
1. *Open big data and analysis platform construction* 2. *Public welfare crowdfunding application and traceability analysis technology* 3. *Intellectual Manufacturing - Platform Optimization* 4. *Data privacy protection technology*	**OS:** Linux, Windows **Distributed Ecosystem:** Hadoop(HDFS, Hive, Hue), Spark **Programming:** Python, Java **Machine Learning Framework:** PySpark(ML)【Decision Trees, K-means】 **IDE:** PyCharm, Eclipse, and Sublime Text 3, Jupyter Notebook **Database:** Hive, MS SQL-Server, MySQL, SQLite **Container:** Docker, Virtual Box, VMware	✧ Data processing with Spark-privacy data de-identification processing, analysis with ML algorithm (Decision Trees, K-means) ✧ Privacy Protection Algorithm Development - Mondrian Algorithm Enhancement ✧ Extension 2. Write as a paper and submit for 2018 IEEE TrustCom(accepted by 2018 IEEE TrustCom) ✧ Standard privacy protection process development [from beginning to end]

	Code Management: GitHub, Bitbucket	✧ Statistical analysis of relevant data using F-test, ANOVA, Regression, chi-square…etc. ✧ Patent application: a k-anonymous method for time series data (accept and issued in 201901 by Taiwan Patent Office) ✧ Privacy Risk Algorithm Development - Developed 3 privacy risk algorithm programs ✧ De-identified data utility analysis algorithm – Developed 4 utility analysis algorithm programs (include 2 ML-related methods)

PROJECTS

2018

Project Name	*Environment*	*Specific Work*
1. *Artificial intelligence collaboration technology and services* 2. *Open innovation collaboration platform* 3. *Data privacy protection technology*	**OS:** Linux, Windows **Distributed Ecosystem:** Hadoop(HDFS, Hive, Hue), Spark **Programming:** Python, Java **Web Platform:** - Application (Jinja2 with Python), Web server (Nginx), DB (PostgreSQL), OS (Linux) - Flask-Redis-Celery **IDE:** PyCharm, Eclipse, and Sublime Text 3, Jupyter Notebook **Database:** Hive, MS SQL-Server, PostgreSQL **Container:** Docker, Virtual Box **Code Management:** Bitbucket	✧ Use the relevant algorithm program developed in 2017 as a web application (Flask-Redis-Celery architecture) ✧ Development Data Automatic Upload Tool (Python) – Allows data to be automatically upload from the backend (with metadata & JSON file) to the CKAN open source big data platform in different ways. ✧ CKAN big data platform, create new platform functions - file preview function (pagination function), permission function (data owner can

		enter this data set as private or public, or freely define allow users) ✧ Data privacy risk analysis/utility analysis module development – extended development of 3 risk analysis algorithm modules and 3 utility analysis algorithm modules

PROJECTS

2019

Project Name	*Environment*	*Specific Work*
1. *AI deep learning value-added platform construction* 2. *Artificial intelligence to create technology and services*	**OS:** Linux, Windows **Programming:** Python, Java, JavaScript, html, CSS **Web Platform:** - Application (Jinja2 with Python), Web server (Nginx), DB (PostgreSQL), OS (Linux) **Machine Learning Framework:** Pytorch(DL) **IDE:** PyCharm, Eclipse, Jupyter Notebook **Database:** PostgreSQL **Container:** Virtual Box **Code Management:** Bitbucket	✧ Privacy Data Synthesis Method Test – Generate synthetic data through GAN and test its availability and risk ✧ CKAN big data platform function development – organization/resource customize columns and inherit the information to other pages ✧ Development data generalization processing API (Java) – Data encryption algorithm (SHA1), address (Chinese regularization), and mask