

# Michael Chin-Chia Yeh

<https://mcyeh.github.io/> ◇ [myeh003@ucr.edu](mailto:myeh003@ucr.edu)

## EDUCATION

---

### University of California, Riverside (UCR)

Ph.D. in Computer Science

### University of California, Los Angeles (UCLA)

M.S. in Mechanical Engineering, Systems and Control

### Virginia Polytechnic Institute and State University (Virginia Tech)

B.S. in Mechanical Engineering

## PROFESSIONAL EXPERIENCE

---

### Visa Research

*Staff Research Scientist*

2018/09 –

Palo Alto, California

- Time Series Data Mining

The time series generated from the transaction database can provide critical insights for payment processing companies. To utilize such information effectively, I have developed machine learning models for time series data under various contexts for transactional time series. The following topics are studied for this project: matrix profile, deep learning, time series forecasting, time series classification/regression, multi-future prediction, multi-model learning (i.e., time series data with graph data), online learning, and anomaly detection.

- Embeddings and Representation Learning

Representation learning is a fundamental building block for analyzing entities in a database. I have conducted researches in three directions for this topic to analyze Visa's data: 1) static embedding learning framework, 2) removing or alleviating the impact of an undesirable feature in the embedding, and 3) scalable and end-to-end embedding learning with downstream models. With the outcomes of these studies, embeddings for entities like merchants, merchant categories, and countries have been learned and adopted for various analytic tasks.

### University of California, Riverside

*Research Assistant*

3 Years

Riverside, California

- All Pairs Similarity Search for Time Series Subsequences (Matrix Profile)

The all-pairs-similarity-search (or similarity join) problem has been extensively studied for text and a handful of other data types. However, there has been little progress on similarity joins for time series subsequences. The goal of the project is to develop an efficient/scalable algorithm solving the similarity join problem for time series data and show the utility of such algorithm when it's treated as a primitive operation for time series. The application of time series join includes visualization, motif/discord mining, clustering, etc.

Supervisor: Prof. Eamonn Keogh

### Research Center for IT Innovation, Academia Sinica

*Research Assistant*

2 Years

Taipei, Taiwan

- Audio Word Representation of Audio Signals

Audio word (AW) representation symbolizes any local audio event as a codeword within a pre-constructed dictionary. Over the course of the project, I have conducted a systematic evaluation with various AW extracting configurations on audio classification/auto-tagging systems. Based on the result of the systematic evaluation, I have proposed a framework that aims to standardize the modularization of the AW representation extraction. I have also examined the possibility of incorporating various ideas (e.g., multi-scale feature learning, bagging) into the AW extraction process to learn better AW representation.

Supervisor: Dr. Yi-Hsuan Yang

## HONORS, AWARDS, AND SERVICE

---

- Honorable mention for 2019 SIGKDD doctoral dissertation award.
- First place in AALTD' 16 time-series classification challenge
- Travel Award: SDM 2018, KDD 2017, ICDM 2016.
- Program Committee: KDD 2019-2022, MiLeTS 2020-2022, AAAI 2021, ECML-PKDD (ADS) 2020.

## PUBLICATION

---

- **Chin-Chia M. Yeh**, Mengting Gu, Yan Zheng, Huiyuan Chen, Javid Ebrahimi, Zhongfang Zhuang, Junpeng Wang, Liang Wang, and Wei Zhang, "Embedding Compression with Hashing for Efficient Representation Learning in Large-Scale Graph," *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2022.
- **Chin-Chia M. Yeh**, Yan Zheng, Junpeng Wang, Huiyuan Chen, Zhongfang Zhuang, Wei Zhang, and Eamonn Keogh, "Error-bounded Approximate Time Series Joins using Compact Dictionary Representations of Time Series," *SIAM International Conference on Data Mining (SDM)*, 2022.
- Huiyuan Chen, **Chin-Chia M. Yeh**, Fei Wang, and Hao Yang, "Graph Neural Transport Networks with Non-local Attentions for Recommender Systems," *ACM Web Conference (WWW)*, 2022.
- Junpeng Wang, Liang Wang, Yan Zheng, **Chin-Chia M. Yeh**, Shubham Jain, and Wei Zhang, "Learning-From-Disagreement: A Model Comparison and Visual Analytics Framework," *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 2022.
- Bo Dong, Yuhang Wu, **Chin-Chia M. Yeh**, Yusan Lin, Yuzhong Chen, Hao Yang, Fei Wang, Wanxin Bai, Krupa Brahmksri, Yimin Zhang, Chinna Kummitha, and Verma Abhisar, "Semi-supervised Context Discovery for Peer-Based Anomaly Detection in Multi-Layer Networks," *International Conference on Information and Communications Security (ICICS)*, 2022.
- **Chin-Chia M. Yeh**, Zhongfang Zhuang, Junpeng Wang, Yan Zheng, Javid Ebrahimi, Ryan Mercer, Liang Wang, and Wei Zhang, "Online Multi-horizon Transaction Metric Estimation with Multi-modal Learning in Payment Networks," *ACM International Conference on Information and Knowledge Management (CIKM)*, 2021.
- Huiyuan Chen, Lan Wang, Yusan Lin, **Chin-Chia M. Yeh**, Fei Wang, and Hao Yang, "Structured Graph Convolutional Networks with Stochastic Masks for Recommender Systems," *International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, 2021.
- Junpeng Wang, Wei Zhang, Hao Yang, **Chin-Chia M. Yeh**, and Liang Wang, "Visual Analytics for RNN-Based Deep Reinforcement Learning," *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 2021.
- Jingzhu He, **Chin-Chia M. Yeh**, Yanhong Wu, Liang Wang, and Wei Zhang, "Mining Anomalies in Subspaces of High-dimensional Time Series for Financial Transactional Data," *Joint European Conference on Machine Learning and Knowledge Discovery in Databases (ECML-PKDD)*, 2021.
- **Chin-Chia M. Yeh**, Zhongfang Zhuang, Yan Zheng, Liang Wang, Junpeng Wang, and Wei Zhang, "Merchant Category Identification Using Credit Card Transactions," *IEEE International Conference on Big Data (BigData)*, 2020.
- **Chin-Chia M. Yeh**, Zhongfang Zhuang, Wei Zhang, and Liang Wang, "Multi-future Merchant Transaction Prediction," *Joint European Conference on Machine Learning and Knowledge Discovery in Databases (ECML-PKDD)*, 2020.
- **Chin-Chia M. Yeh**, Dhruv Gelda, Zhongfang Zhuang, Yan Zheng, Liang Gou, and Wei Zhang, "Towards a Flexible Embedding Learning Framework," *IEEE International Conference on Data Mining Workshop on Multi-Source Data Mining (MSDM)*, 2020.

- Yan Zhu, **Chin-Chia M. Yeh**, Zachary Zimmerman, and Eamonn Keogh, “Matrix Profile XVII: Indexing the Matrix Profile to Allow Arbitrary Range Queries,” *IEEE International Conference on Data Engineering (ICDE)*, 2020.
- Zhongfang Zhuang, **Chin-Chia M. Yeh**, Liang Wang, Wei Zhang, and Junpeng Wang, “Multi-stream RNN for Merchant Transaction Prediction,” *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining Workshop on Machine Learning in Finance (MLF)*, 2020.
- **Chin-Chia M. Yeh**, Yan Zhu, Hoang Anh Dau, Amirali Darvishzadeh, Mikhail Noskov, and Eamonn Keogh, “Online Amnestic DTW to allow Real-Time Golden Batch Monitoring,” *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2019.
- Shaghayegh Gharghabi, **Chin-Chia M. Yeh**, Yifei Ding, Wei Ding, Paul Hibbing, Samuel LaMunion, Andrew Kaplan, Scott E. Crouter, and Eamonn Keogh, “Domain Agnostic Online Semantic Segmentation for Multi-dimensional Time Series,” *Data Mining and Knowledge Discovery (DMKD)*, 2019.
- **Chin-Chia M. Yeh**, “Towards a Near Universal Time Series Data Mining Tool: Introducing the Matrix Profile,” University of California, Riverside, 2018.
- Alireza Abdoli, Amy C. Murillo, **Chin-Chia M. Yeh**, Alec C. Gerry, and Eamonn J. Keogh, “Time Series Classification to Improve Poultry Welfare,” *IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2018.
- Nader S. Senobari, Gareth J. Funning, Eamonn Keogh, Yan Zhu, **Chin-Chia M. Yeh**, Zachary Zimmerman, and Abdullah Mueen, “Super-Efficient Cross-Correlation (SEC-C): A Fast Matched Filtering Code Suitable for Desktop Computers,” *Seismological Research Letters*, 2018.
- Yan Zhu, **Chin-Chia M. Yeh**, Zachary Zimmerman, Kaveh Kamgar, and Eamonn Keogh, “Matrix Profile XI: SCRIMP++: Time Series Motif Discovery at Interactive Speeds,” *IEEE International Conference on Data Mining (ICDM)*, 2018.
- Yan Zhu, Zachary Zimmerman, Nader S. Senobari, **Chin-Chia M. Yeh**, Gareth Funning, Abdullah Mueen, Philip Brisk, and Eamonn Keogh, “Exploiting a Novel Algorithm and GPUs to Break the Ten Quadrillion Pairwise Comparisons Barrier for Time Series Motifs and Joins,” *Knowledge and Information Systems (KIS)*, 2018.
- **Chin-Chia M. Yeh**, Yan Zhu, Liudmila Ulanova, Nurjahan Begum, Yifei Ding, Hoang Anh Dau, Zachary Zimmerman, Diego F. Silva, Abdullah Mueen, and Eamonn Keogh, “Time Series Joins, Motifs, Discords and Shapelets: a Unifying View that Exploits the Matrix Profile,” *Data Mining and Knowledge Discovery (DMKD)*, 2018.
- Shaghayegh Gharghabi, Yifei Ding, **Chin-Chia M. Yeh**, Kaveh Kamgar, Liudmila Ulanova, and Eamonn Keogh, “Matrix Profile VIII: Domain Agnostic Online Semantic Segmentation at Superhuman Performance Levels,” *IEEE International Conference on Data Mining (ICDM)*, 2017.
- **Chin-Chia M. Yeh**, Nickolas Kavantzaz, and Eamonn Keogh, “Matrix Profile VI: Meaningful Multidimensional Motif Discovery,” *IEEE International Conference on Data Mining (ICDM)*, 2017.
- **Chin-Chia M. Yeh**, Nickolas Kavantzaz, and Eamonn Keogh, “Matrix Profile IV: Using Weakly Labeled Time Series to Predict Outcomes,” *Proceedings of the VLDB Endowment (VLDB)*, 2017.
- **Chin-Chia M. Yeh**, Helga Van Herle, and Eamonn Keogh, “Matrix Profile III: The Matrix Profile Allows Visualization of Salient Subsequences in Massive Time Series,” *IEEE International Conference on Data Mining (ICDM)*, 2016.
- Yan Zhu, Zachary Zimmerman, Nader S. Senobari, **Chin-Chia M. Yeh**, Gareth Funning, Abdullah Mueen, Philip Brisk, and Eamonn Keogh, “Matrix Profile II: Exploiting a Novel Algorithm and GPUs to Break the One Hundred Million Barrier for Time Series Motifs and Joins,” *IEEE International Conference on Data Mining (ICDM)*, 2016.

- **Chin-Chia M. Yeh**, Yan Zhu, Liudmila Ulanova, Nurjahan Begum, Yifei Ding, Hoang Anh Dau, Diego F. Silva, Abdullah Mueen, and Eamonn Keogh, “Matrix Profile I: All Pairs Similarity Joins for Time Series: A Unifying View that Includes Motifs, Discords and Shapelets,” *IEEE International Conference on Data Mining (ICDM)*, 2016.
- Diego F. Silva, **Chin-Chia M. Yeh**, Gustavo E. A. P. A. Batista, Eamonn Keogh, “SiMPle: Assessing Music Similarity Using Subsequences Joins,” *International Society for Music Information Retrieval Conference (ISMIR)*, 2016.
- **Chin-Chia M. Yeh**, Ping-Keng Jao, and Yi-Hsuan Yang. *The AWtoolbox for characterizing audio information*, Academia Sinica, Technical Report, 2015.
- Li Su, **Chin-Chia M. Yeh**, Jen-Yu Liu, Ju-Chiang Wang, and Yi-Hsuan Yang, “A Systematic Evaluation of the Bag-of-frames Representation for Music Information Retrieval,” *IEEE Transactions Multimedia (TMM)*, 2014.
- **Chin-Chia M. Yeh**, Ping-Keng Jao, and Yi-Hsuan Yang. “AWtoolbox: Characterizing Audio Information Using Audio Words,” *ACM International Conference Multimedia (MM)*, 2014.
- **Chin-Chia M. Yeh**, Ju-Chiang Wang, Yi-Hsuan Yang, and Hsin-Min Wang, “Improving Music Auto-tagging by Intra-song Instance Bagging,” *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014.
- Ping-Keng Jao, **Chin-Chia M. Yeh**, and Yi-Hsuan Yang, “Modified LASSO Screening for Audio Word-based Music Classification Using Large-scale Dictionary,” *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014.
- **Chin-Chia M. Yeh** and Yi-Hsuan Yang, “Towards a More Efficient Sparse Coding Based Audio-word Feature Extraction System,” *Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2013.
- **Chin-Chia M. Yeh**, Li Su, and Yi-Hsuan Yang, “Dual-layer Bag-of-frames Model for Music Genre Classification,” *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013.
- Jen-Yu Liu, **Chin-Chia M. Yeh**, Yuan-Ching Teng, and Yi-Hsuan Yang, “Bilingual Analysis of Song Lyrics and Audio Words,” *ACM International Conference Multimedia (MM)*, 2012.
- **Chin-Chia M. Yeh** and Yi-Hsuan Yang, “Supervised Dictionary Learning for Music Genre Classification,” *ACM International Conference on Multimedia Retrieval (ICMR)*, 2012.

## PATENT

---

- **Michael Yeh**, Liang Gou, Wei Zhang, Dhruv Gelda, Zhongfang Zhuang, Yan Zheng, “System, Method, and Computer Program Product for Analyzing a Relational Database Using Embedding Learning,” US Patent App. 17/066,852.
- Yan Zheng, Yuwei Wang, Wei Zhang, **Michael Yeh**, and Liang Wang, “Unsupervised Embeddings Disentanglement Using a GAN for Merchant Recommendations,” US Patent App. 16/688,847.
- Zhongfang Zhuang, **Michael Yeh**, Liang Wang, Wei Zhang, and Junpeng Wang, “System, Method, and Computer Program Product for Multivariate Event Prediction Using Multi-Stream Recurrent Neural Networks,” US Patent App. 17/148,984.

## RELEVANT SKILL

---

### *Programming Language*

- Proficient: MATLAB, Python, and L<sup>A</sup>T<sub>E</sub>X
- Familiar: Java, C#, and C++

### *Natural Language*

- Bilingual Proficiency: English, Mandarin Chinese