Michael Chin-Chia Yeh

https://mcyeh.github.io/ ◇ 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 2018/09 –
Staff Research Scientist Palo Alto, California

- Time Series Predictive Model

  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. Topics include deep learning, time series forecasting, time series classification/regression, multi-future prediction, multi-model learning (i.e., time series data with relationship data), online learning, and anomaly detection are studied for this project.
- 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 in Visa.

University of California, Riverside 2015/07 – 2018/07
Research Assistant 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, classification, and representation learning.
  Supervisor: Prof. Eamonn Keogh

Research Center for IT Innovation, Academia Sinica 2011/07 – 2012/06, 2013/07 – 2014/08
Research Assistant Taipei, Taiwan

- Audio Word Representation of Audio Signals

  Audio word (AW) representation is characteristic of its ability of symbolizing any local audio event as a codeword within a pre-constructed dictionary. Over the course of the project, I have conducted systematic evaluation with various AW extracting configuration on audio classification and auto-tagging systems, proposed a framework aims to standardize the modularization of the AW representation extraction, and developed an open-source graphic user interface application intended to facilitate the implementation and development of various AW extraction procedural under aforementioned framework. On top of that, in order to improve the effectiveness of AW representation, I have examined the possibility of incorporating various ideas (e.g., multi-scale feature learning, bagging) into the AW extraction process.
  Supervisor: Dr. Yi-Hsuan Yang

OTHER EXPERIENCE

Bosch Research and Technology Center 2017/06 - 2017/09
Research Intern Palo Alto, California

Aspen Technology 2016/06 - 2016/09
Data Scientist Intern Bedford, Massachusetts

PUBLICATION

- **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 Int. on Conf. 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," *Int. ACM SIGIR Conf. 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 Conf. 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 Int. Conf. on Big Data* (BigData), 2020.
- **Chin-Chia M. Yeh**, Zhongfang Zhuang, Wei Zhang, and Liang Wang, "Multi-future Merchant Transaction Prediction," *Joint European Conf. 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 Int. Conf. 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 Int. Conf. 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 Int. Conf. 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 Int. Conf. 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.
- Alireza Abdoli, Amy C. Murillo, **Chin-Chia M. Yeh**, Alec C. Gerry, and Eamonn J. Keogh, "Time Series Classification to Improve Poultry Welfare," *IEEE Int. Conf. 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 Int. Conf. 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 Int. Conf. on Data Mining* (ICDM), 2017.
- **Chin-Chia M. Yeh**, Nickolas Kavantzas, and Eamonn Keogh, "Matrix Profile VI: Meaningful Multidimensional Motif Discovery," *IEEE Int. Conf. on Data Mining* (ICDM), 2017.
- **Chin-Chia M. Yeh**, Nickolas Kavantzas, 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 Int. Conf. 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 Int. Conf. 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 Int. Conf. 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," *Int. Society for Music Information Retrieval Conf.* (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 Trans. Multimedia* (TMM), 2014.
- **Chin-Chia M. Yeh**, Ping-Keng Jao, and Yi-Hsuan Yang. "AWtoolbox: Characterizing Audio Information Using Audio Words," *ACM Int. Conf. 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 Int. Conf. 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 Int. Conf. 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 Int. Conf. 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 Int. Conf. Multimedia* (MM), 2012.
- **Chin-Chia M. Yeh** and Yi-Hsuan Yang, "Supervised Dictionary Learning for Music Genre Classification," *ACM Int. Conf. on Multimedia Retrieval* (ICMR), 2012.

RELEVANT SKILL

Programming Language

- Proficient: MATLAB, Python, and LaTeX
- Familiar: Java, C#, and C++

Language

- Mandarin Chinese (mother tongue)