JINGWEI ZUO

⌂ Versailles, France
📞 (+33) (0)6 50 13 98 29
@ jingwei.zuo.career@gmail.com
🌐 jingweizuo.com   in linkedin.com/in/jingweizuo   ⚙ github.com/JingweiZuo

As a new-graduate PhD in computer science with hands-on experience in machine (deep) learning, time series analysis and data science ecosystems, I'm open to **Data Scientist, Machine Learning Engineer** and **Machine Learning Researcher** roles.

EDUCATION

Ph.D. in Machine Learning — Oct. 2018 - May 2022
Université Paris-Saclay, advisors: Prof. Karine Zeitouni, Dr. Yehia Taher — *Versailles, France*
- **Thesis**: Representation Learning and Forecasting for Inter-related Time Series

Master in Data Science, graduated with highest honors — Sept. 2017 - Sept. 2018
Université Paris-Saclay — *Versailles, France*

Engineer in Electronics and Computer Sciences — Sept. 2014 - Aug. 2017
Polytech Sorbonne - Sorbonne Université — *Paris, France*

Bachelor of Science in Optical and Electronic Information — Sept. 2011 - Aug. 2014
Huazhong University of Science and Technology (HUST) — *Wuhan, China*

PROFESSIONAL EXPERIENCES

Research Intern, DAVID Lab - University of Versailles — Apr. 2018 - Sept. 2018
Topic: "Big Time Series Classification ⚙" — *Versailles, France*
- Utilized **Python** and **Scala** to implement supervised machine learning algorithms for time series activity recognition on real artwork transportation data, which is validated by our collaborators in C2RMF.
- Proposed a distributed algorithm on **Spark** for time series classification, which reduced total processing time by 75% in local Docker cluster, by 99% in AWS EMR cluster with 30 distributed nodes.
- **Publication**: **J. Zuo**, K. Zeitouni and Y. Taher. Exploring Interpretable Features for Large Time Series with SE4TeC, demo, International Conference on Extending Database Technology (EDBT 2019), *CORE ranking*: A

Part-time R&D Intern, PETRUS Team - Inria Saclay — Dec. 2017 - Mar. 2018
Topic: "Mobile Personal Medical Folder on PlugDB" — *Versailles, France*
- Utilized **Java** (Android) and **SQL** to implement a mobile personal medical folder on a secure personal database (PlugDB).
- This project is set to be deployed in the Central Hospital of Yaoundé for over 1000 tracked patients with diabetes.

Software Engineer Intern, Orange Labs — Mar. 2017 - Aug. 2017
Topic: "Authentication Technologies on IoT devices" — *Lannion, France*
- Utilized **C** and **C#** to implement authentication technologies on IoT devices.

SKILLS

- **Programming Languages**: Python, Java, Scala, C, C++, C#, R, JavaScript, SQL, Matlab

- **Big Data & Machine Learning**: Spark, Hadoop, Kafka, Redis, InfluxDB, MongoDB, PostgreSQL, ElasticSearch, Kibana; scikit-learn, Tensorflow, Keras, PyTorch

- **Miscellaneous Technologies**: A/B testing, ETL, Data science pipeline (cleansing, wrangling, visualization, modeling, interpretation), Statistics, Experimental design, Hypothesis testing, Git, Docker, Kubernetes, AWS

- **Languages**: Chinese (Native); French, English (Fluent).

PROJECTS

- **Supervised Learning on Time Series Stream ⚙**: *Dynamic Representation Learning for Classification*
  - Utilized **Python** and **Flask** to implement scalable algorithms and a visualization web application for dynamic representation learning from time series stream.
  - Achieved the best classification results on 12 UCR time series datasets and saved up to 96% of processing time.
  - **Publication**: **J. Zuo** et al., Incremental and Adaptive Feature Exploration over Time Series Stream. IEEE International Conference on Big Data (IEEE BigData 2019), acceptance rate: 19%
  - **Publication**: **J. Zuo** et al., ISETS: Incremental Shapelet Extraction from Time Series Stream, demo, ECML-PKDD'19, CORE ranking: A

- **Streamer ⚙**: *Open-Source Framework for Continuous Learning in Data Streams*
  - Utilized **Java** to integrate the time series stream algorithms into a practical data stream learning pipeline: data stream ingestion (Kafka), data storage (InfluxDB), model storage (Redis) and visualisation (Kibana).
  - Enriched the functionalities of Streamer with our collaborators in CEA-List.

- **Semi-supervised Learning on Time Series ⚙**: Interpretable Representation Learning
  - Utilized **Keras** to implement an autoencoder-based model, which tackles the label shortage problem in multivariate time series learning.
  - Achieved the best accuracy over the state-of-the-art approaches on the UEA Archive containing 30 multivariate time series datasets, while guaranteeing the model's interpretability.
  - **Publication**: **J. Zuo** et al., SMATE: Semi-supervised Spatio-Temporal Representation Learning on Multivariate Time Series, IEEE International Conference on Data Mining (ICDM 2021), *CORE ranking*: A*

- **Traffic Forecasting**: *Incomplete Time Series Forecasting with Graph Neural Networks*
  - Utilized **PyTorch** on IDRIS GPU cluster to test and implement time series forecasting models for real-life traffic data, which contains over 16 million observations with complex missing values.
  - Achieved better predictions than both the statistic models (e.g., ARIMA) and recent deep learning approaches.
  - **Publication**: **J. Zuo** et al., Graph Convolutional Networks for Traffic Forecasting with Missing Values. under review by DMKD, Springer.

- **PolluScope 🌐**: *Distributed Query, Multi-view Learning, Pollution Activity Recognition, Classification*
  - Combined **PySpark** and **Scala**-**UDF** to implement distributed query operators on multi-granular time series database, which reduced the query's time by 90% compared to PySpark.
  - Utilized **Python** and **scikit-learn** to implement a multi-view learning model from the multi-source time series, which improved the accuracy by 9% over single-view learners on the actual pollution exposure data collected in Île-de-France.
  - **Publication**: H. El Hafyani, M. Abboud, **J. Zuo**, et al., Learning the Micro-environment from Rich Trajectories in the context of Mobile Crowd Sensing - Application to Air Quality Monitoring, under review by Geoinformatica, Springer
  - **Publication**: H. El Hafyani, M. Abboud, **J. Zuo** et al., Tell Me What Air You Breathe, I Tell You Where You Are, demo, International Symposium on Spatial and Temporal Databases (SSTD 2021)
  - **Publication**: M. Abboud, H. El Hafyani, **J. Zuo** et al., Micro-environment Recognition in the context of Environmental Crowdsensing, 4th workshop on Big Mobility Data Analytics (BMDA@EDBT 2021)

COURSEWORK

- **Graduate**: Data Science (Database, Data Integration, Data Mining, Machine learning, Cloud Computing), Programming (Java, Scala, C++, Software Engineering), Signal Processing, Project Management, Entrepreneurship, etc.
- **Undergraduate**: Mathematics, Physics, Computer Architecture, Operating Systems, etc.
- **5th Int. winter school on Big Data**: one-week training program on Big Data at University of Cambridge
- **3rd DS^3 summer school**: one-week training program on Data Science at École Polytechnique
- **IA^2 autumn school 2019**: one-week training program on Artificial Intelligence at INSA Lyon

PROFESSIONAL SERVICES AND AWARDS

- **Teaching assistant at Univ. of Versailles**: 2019-2021 (128 hours), for undergraduate and graduate level courses on Data Mining, Database, Object-oriented Programming (Python, Java), Software Engineering
- **Organizing committee**: Web chair of IEEE MDM'20, committee member of JDSE'19
- **Program committee**: ECML-PKDD'20
- **Reviewing activities**: AAAI'21, IEEE ICDE'21, SSTD'21, IEEE DSAA'20
- **Student travel award & volunteering**: IEEE International Conference on Big Data 2019