Application Form - Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship Applicants

Applicant Details

Applicant Name: Xinyu Chen

Applicant Email: chenxy346@gmail.com

Applicant Institution: University of Montreal

Applicant Mailing Address: 22-3195 Ellendale Avenue, Montreal, Canada

PhD Completion Date / Expected Date: 12/20/2023

Citizenship

Canadian Citizen ☐

Permanent Resident ☐

Indicate date of landing as per Form IMM 1000, if applicable: Click or tap here to enter text.

Foreign Student with valid student Visa for the full work term ☒

Supervisor(s) Information

Supervisor Name: Eric Miller

Supervisor Email: eric.miller@utoronto.ca

Supervisor Primary Institutional Affiliation: University of Toronto

Supervisor's University of Toronto Home Division: Faculty of Applied Science & Engineering

Supervisor University of Toronto Home Academic Unit: Department of Civil & Mineral Engineering

Co-supervisor, if applicable –

Co-supervisor Name: Click or tap here to enter text.

Co-supervisor Email: Click or tap here to enter text.

Co-Supervisor Primary Institutional Affiliation: Click or tap here to enter text.

Co-Supervisor's University of Toronto Home Division: Click or tap here to enter text.

Co-Supervisor University of Toronto Home Academic Unit: Click or tap here to enter text.

Disclaimer

Do you currently hold a postdoctoral fellowship?

Yes ☐ No ☒

For complete program guidelines, please refer to the program page, here: https://schmidtfellows.utoronto.ca/apply

Narrative Sections of the Application (Required)

Presentation Standards

- Body text minimum 12 pt. font (e.g., Times New Roman), black type
- Single-spaced, with no more than 6 lines of type per inch
- All margins set at a minimum of 3/4" (1.87 cm)
- Do not use condensed/narrow font sizes, type density, or line spacing. Smaller text in tables, charts, figures, and graphs is acceptable, if it is legible when the page is viewed at 100%.

1. Research Proposal (3 pages maximum)

Describe the proposed research project, highlighting the following: 1) Project title and rationale, 2) Explanation of how the candidate plans to use AI to transform their research, 3) Research objectives, 4) Research approach and plans to engage in research that spans across traditional disciplinary boundaries, 5) Feasibility and timeline, and 6) Anticipated Impact.

2. Bibliography (1 page maximum)

Please provide a list of citations for all works referenced in the research proposal, formatted in agreement with your discipline standards. All citations should be clear and complete.

3. Training Plan (1 page maximum)

Please outline your current knowledge of AI-supported research methodologies and what you will need to learn to make your research project successful. If you have identified specific training opportunities you will pursue (i.e., courses, workshops, seminars), please list them here.

4. Candidate Statement (2 pages maximum)

Please address the following in your candidate statement: 1) Your rationale for applying to the Schmidt AI in Science Fellowship Program, 2) 1-2 of your most significant research contributions (e.g., publications, presentations, research collaborations, prestigious research recognitions, intellectual property), 3) How this fellowship aligns with your career development and future goals, and 4) Your commitment to equity, diversity, and inclusion. Recommended guidelines are half a page per point.

5. Significance of Leadership Contributions (1 page maximum)

Please identify 2-3 leadership experiences which you deem to be most impactful and relevant to this fellowship. For each example, describe its significance in terms of your leadership and impact on your discipline, institution, and society (if applicable). Also comment on how these experiences contribute to your career goals. Examples of leadership may include, but are not limited to teaching activities, supervisory activities, advisory and administrative activities, community and volunteer activities, and memberships.

6. Applicant's Curriculum Vitae

For complete program guidelines, please refer to the program page, here:
https://schmidtfellows.utoronto.ca/apply

7. Applicant Demographic Survey*

A reminder that candidates are also required to complete and submit the [Applicant Demographic Survey] as part of their application.

***Notice of Collection:** The following short survey is designed for applicants to the Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship program at the University of Toronto. Gathering demographic data supports our ability to evaluate the diversity of our pool of applicants in each competition and allows us to contextualize our adjudications processes and inform our Equity, Diversity, and Inclusion reporting and best practices.

While completion of the survey is a required component of the application, candidates have the option of answering "Prefer not to reply" for each of the questions.

The information collected via this form will be kept confidential and will not be used as part of the evaluation of your application or any other future application. Personal information you provide the University is at all times protected in accordance with the Freedom of Information and Protection of Privacy Act. Data collected via this survey will be held securely and stored separately from all other employment or personal data. This information will be used solely for the purpose of understanding the diversity of our applicant pools and may not be accessed or used for other purposes.

The data will be held in the office of the Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship program at the University of Toronto and will be accessible only to two designated individuals: the Schmidt Fellows Program Manager and the Director of Equity, Diversity and Inclusion in A&S. Only aggregated and deidentified data will be shared and reported internally and with the program funder as part of our administration and reporting processes; no record-level or identifying information will be shared or reported.

For more information, please contact us at schmidtfutures@utoronto.ca.

Understanding Urban Mobility and Travel Modes with Machine Learning

Background: In the past decades, urbanization and digitization have stimulated urban system evolution and transport mode development (Bettencourt, 2021). We also witnessed a large number of innovative attempts that reshape urban transport systems and make the city "smart" and sustainable, including emerging travel modes (e.g., ridesharing, bikesharing, e-bike, and e-scooter), emerging technologies (e.g., electric vehicle and autonomous vehicle), and emerging infrastructure (e.g., cycling and walking facilities). These inspire communities to adopt behaviours in favour of sustainable development. However, these also complicate the choice of travel modes and the behaviour of urban mobility. Furthermore, considering the great change of human mobility behaviours caused by COVID-19, it is meaningful to investigate urban mobility and travel mode transition. To this end, this project aims at understanding urban mobility behaviours and patterns from individual travel data, in which research results are of great significance in the process of urban planning, development, and policy-making.

Nowadays, considerable amounts of human mobility data with various travel modes are collected through advanced information systems on a (nearly) real-time basis, providing tremendous opportunities for understanding complicated characteristics of urban mobility and revealing spatiotemporal patterns. These data include public transit data (Pelletier et al., 2011), taxi trace data (Zheng, 2015), and smartphone trace data (Harding et al., 2021; Vaughan et al., 2020) to name but a few. To gain insight into urban transport systems from researchers and practitioners, many cities around the world also have started to make human mobility data publicly available, while trips are anonymized or aggregated. For example, New York City (NYC) provides a catalog of public datasets, including bikesharing trip data[1], taxi trip data[2], and other mobility related data[3]. In terms of ridesharing, Uber Movement project shares anonymized data (e.g., movement speeds, travel times, and mobility heatmap) aggregated from over ten billion trips to help urban planning around the world.[4] In terms of bikesharing, the data of many cities in North America are publicly available (e.g., BIXI Montreal[5] and Bike Share Toronto[6]) and they follow a standardized data feed specification. These mobility data allow one to discover human mobility behaviours, identify travel modes, and improve the quality of transport systems. In the literature, Artificial Intelligence (AI), relying on computer modelling and programming (Goodfellow et al., 2016), has demonstrated great potential and applications to urban studies

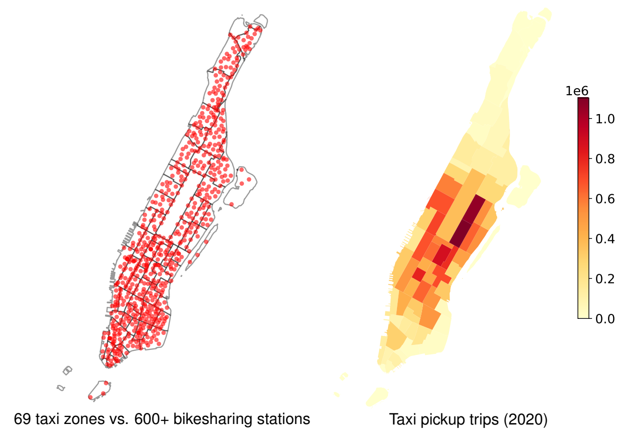

Figure 1: Visualization of urban human mobility with two travel modes (i.e., taxi and bikesharing) in Manhattan, NYC, USA.

[1]Citi Bike trip data are available at https://citibikenyc.com/system-data, starting from June 2013 until now.

[2]NYC taxi trip data are available at https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page, starting from 2009 until now.

[3]The New York City Department of Transportation (NYC DOT) Open Data feeds are available at https://www.nyc.gov/html/dot/html/about/datafeeds.shtml.

[4]Uber movement data are available at https://movement.uber.com/.

[5]BIXI Bike trip data are available at https://bixi.com/en/open-data-2/, starting from 2014 until now.

[6]Bike Share Toronto trip data are available at https://open.toronto.ca/dataset/bike-share-toronto-ridership-data/.

(Zheng, 2019). But in the data modelling process, one is still required to reformulate human mobility problems with machine learning algorithms.

Scientific Questions & Objectives: Due to the availability of big mobility data and the development of machine learning algorithms, there is a great opportunity for reformulating human mobility problems from a data-driven perspective. Since mobility data show complicated data behaviours and strong domain-specific knowledge, the primary questions arise with respect to: 1) How to represent the multi-dimensional mobility data? 2) How to project the mobility of various travel modes onto the same data space for analysis? 3) How to fuse different sources of mobility data? Essentially, in a given time slot, each mobility record can be represented as the movement started from an origin location to a destination location. Accordingly, aggregating these mobility records would lead to the notion of an origin-destination matrix. In a dynamical system, these origin-destination matrices can be further referred to as tensors that are associated with origin, destination, time slot, and travel mode dimensions. This project aims at discovering dynamic patterns from human mobility, identifying travel modes from individual trip data, and forecasting mobility with various travel modes simultaneously. More concretely, the objectives will include:

- (**Objective A: Mobility Pattern Discovery**) Human mobility data may illuminate how the essential mobility patterns behave, yet a substantial barrier remains: How to characterize these data and reveal the underlying patterns? This requires us to develop a well-suited machine learning framework to characterize the time-varying and nonlinear system behaviour of long-term urban mobility. In particular, the model is expected to identify how the system evolves (e.g., mobility reduction[7] and travel mode transition) caused by COVID-19 and characterize the evolution of urban mobility with multi-source data (e.g., population, mobility, land-use, and economic activities). In addition, as COVID-19 dramatically changed the human mobility behaviour, one is required to find a unified framework to reveal the mobility transition and quantify the movement efficiency of transport networks (e.g., traffic congestion) in multi-scale urban systems during past 5-10 years.

- (**Objective B: Travel Mode Identification**) Human mobility can be categorized according to travel mode, including active travel (e.g., walking and cycling), automobile, and public transit (e.g., metro and bus). As cellular network base transceiver stations can gather detailed individual trips through smartphones, automatically identifying or detecting travel modes of these individual trips is important for understanding urban mobility. However, there are some technical challenges: 1) Smartphone trace data do not show the ground-truth travel mode of individual trips, and thus the classification task is one of unsupervised learning. 2) If public transit data and travel survey data are available, then how to develop a data fusion framework and utilize these data as the ground-truth to impute travel modes of smartphone traces? These challenges require us to develop an appropriate supervised learning algorithm. In particular, the algorithm is expected to increase the reliability of travel mode identification and eliminate model biases with limited ground-truth data (e.g., travel survey data[8,9]).

- (**Objective C: Mobility Forecasting**) Large-scale and multi-mode mobility data are becoming ubiquitous in urban transport systems. Making accurate prediction on these

[7] Riderships of US public transit in 2021 and 2022 are far from the peak at the time before COVID-19, see the weekly ridership chart at `https://transitapp.com/APTA`.

[8] Transportation Tomorrow Survey reports: `http://dmg.utoronto.ca/transportation-tomorrow-survey/tts-reports/`.

[9] THATS (Toronto Household Activity-Travel Survey) application for surveying travel data: `https://tmg.utoronto.ca/thats/`.

data has become a critical challenge due to not only high dimensionality but also the mobility conservation of traffic flow with different travel modes. In this case, the task is to dynamically forecast short-/long-term urban mobility with different travel modes simultaneously and incorporate the flow conservation of urban mobility in the transport system. This requires us to address 1) the inconsistency of spatial resolutions between different travel modes (e.g., 69 taxi zones vs. 600+ bikesharing stations in Manhattan, see Fig. 1), and 2) the inconsistency between demand and realistic traffic/trip volume.

Methodology: The goal of this project is to develop supervised/unsupervised learning algorithms for analyzing human mobility data. Concretely, the methodologies will include:

- (**Objective A: Mobility Pattern Discovery**) In this case, we plan to make full use of long-term mobility data with various travel modes (e.g., taxi, ridesharing, public transit, and bike) in urban areas. In the modelling process, we first assume that long-term human mobility can be characterized as a linear dynamical system and build a time-varying tensor autoregression for discovering dynamic patterns from mobility data. In the nonlinear case, we plan to integrate deep neural networks into the autoregression process. Empirically, we plan to utilize city-wide traffic state data (e.g., travel times of NYC taxi trip data) and develop a non-negative tensor factorization algorithm for pattern discovery. In the meanwhile, we plan to find the evidence of travel mode transition of urban mobility, e.g., preference of active travel modes in the post-COVID-19 era.
- (**Objective B: Travel Mode Identification**) As smartphone trace data, traffic survey data, and public transit data are important sources for sensing human mobility, we plan to develop a data fusion framework and explore deep neural networks to improve the tool for imputing/detecting travel modes (e.g., active modes, automobile, and public transit) of smartphone traces. In the modelling process, to increase the reliability of travel mode detection, we plan to supplement the training set with bikesharing trip data. To eliminate the model bias due to the limited ground-truth data, we also plan to generate trip samples through deep generative models (e.g., diffusion model).
- (**Objective C: Mobility Forecasting**) We plan to develop deep learning based forecasting models for hierarchical mobility with different travel modes. The forecasts of lower spatial resolution (e.g., neighbourhood) are allocated to the higher spatial resolution (e.g., station) according to flow conservation and demand-supply considerations.

Timeline: The project feasibility and timeline are summarized as follows.

Tasks	2024		2025			2026
	Summer	Fall	Winter	Summer	Fall	Winter
Literature review	■	■	■			
Training (e.g., courses/seminars)	■	■	■			
Objective A		■	■			
Objective B			■	■	■	
Objective C				■	■	■
Papers & technical reports			■	■	■	■

Anticipated Impact: Understanding urban human mobility shows great potential for pursuing efficient transportation management and supporting social and economic activities. In this project, building theoretical and empirical frameworks, developing machine learning algorithms, and designing scientific computing techniques would support data-driven policy-making for urban development and provide insight into AI's applications to urban science.

References

Bettencourt, L. M. (2021). Introduction to urban science: evidence and theory of cities as complex systems.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press.

Harding, C., Faghih Imani, A., Srikukenthiran, S., Miller, E. J., and Nurul Habib, K. (2021). Are we there yet? assessing smartphone apps as full-fledged tools for activity-travel surveys. *Transportation*, 48:2433–2460.

Pelletier, M.-P., Trépanier, M., and Morency, C. (2011). Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies*, 19(4):557–568.

Vaughan, J., Imani, A. F., Yusuf, B., and Miller, E. J. (2020). Modelling cellphone trace travel mode with neural networks using transit smartcard and home interview survey data. *European Journal of Transport and Infrastructure Research*, 20(4):269–285.

Zheng, Y. (2015). Trajectory data mining: an overview. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 6(3):1–41.

Zheng, Y. (2019). *Urban computing*. MIT Press.

Training Plan

My current knowledge of AI-supported research methodologies include:

- Convex/nonconvex optimization
- Matrix/tensor computations
- Low-rank regression/autoregression
- Matrix/tensor factorization
- Bayesian inference
- Statistical learning
- Gaussian process
- Convolution & Fourier transform
- Recurrent neural networks
- Generative adversarial networks
- Diffusion models
- Reinforcement learning

To ensure the successful implementation of the proposed project, I need to learn the following AI algorithms:

- Convolutional neural networks
- Residual networks
- Graph neural networks
- Attention models

I plan to gain a better understanding of AI algorithms and machine learning by taking the following Computer Science courses at the University of Toronto:

- CSC311H1 - Introduction to Machine Learning
- CSC413H1 - Neural Networks and Deep Learning
- CSC466H1 - Numerical Methods for Optimization Problems

In addition, I will also attend some machine learning and data science-related seminars and workshops at the University of Toronto Data Sciences Institute.

Candidate Statement

Dear program committee,

I am Xinyu Chen, a PhD candidate in Civil Engineering (Transportation) at the University of Montreal. My PhD thesis is titled "Matrix and Tensor Models for Spatiotemporal Traffic Data Imputation and Forecasting" and is supported by the Institute for Data Valorization (IVADO) PhD Excellence Scholarship with a value of $100,000. I am expected to graduate in the Fall of 2023. I am a highly self-motivated and independent researcher with a strong interest in applying for the University of Toronto's Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship. First, I possess a solid interdisciplinary background in Data Science, Machine Learning, Transportation Science, and Urban Science. Throughout my academic journey, I have consistently demonstrated a strong motivation towards developing both theoretical and empirical frameworks, creating machine learning algorithms, tackling complex optimization problems, and designing scientific computing techniques. During my PhD studies, I have developed a deep understanding of urban science and transport systems from a data-driven perspective. I have worked extensively with diverse mobility datasets, including taxi/ridesharing/bikesharing trip data collected from cities such as NYC, Seattle, and Montreal, taxi trajectory data, and individual trip data from metro systems. I have also formulated theoretical frameworks for analyzing individual trip behavior and mobility patterns. Second, I believe that this fellowship offers tremendous opportunities to advance AI techniques and explore cutting-edge applications in urban science. Moreover, the University of Toronto Transportation Research Institute (UTTRI) provides an ideal environment for investigating human mobility data and developing urban transport models. Consequently, I am eager to pursue the proposed research project titled "Understanding Human Mobility and Travel Modes with Machine Learning." This project encompasses tasks related to mobility pattern discovery, travel mode identification, and mobility forecasting. It is expected to yield valuable insights in both the field of urban science and the development of AI algorithms.

In addition to the above, I have made algorithmic and methodological contributions to the fields of AI & Machine Learning, Data Science, and Transportation Science. I have a strong track record in publishing as the first author in top-tier Computer Science & Transportation Science journals, including *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *IEEE Transactions on Knowledge and Data Engineering*, *Transportation Research Part C: Emerging Technologies*, and *IEEE Transactions on Intelligent Transportation Systems*. One recent work addressed the challenge of multidimensional time series forecasting (e.g., spatiotemporal traffic forecasting) in the presence of missing data.[1] This study presented a Bayesian temporal factorization framework designed to characterize both low-rankness property and temporal correlations of time series data, addressing the issues of high-dimensionality and sparsity. Another recent work addressed the problem of discovering interpretable dynamic patterns from spatiotemporal data.[2] The proposed time-varying autoregression model with tensor factorization can identify dynamic patterns in time-varying systems, such as capturing the dynamic pattern changes in urban mobility due to COVID-19. The model lays an insightful foundation for understanding complex spatiotemporal data in real-world dynamical systems. In addition to these studies, I have made significant contributions to an open-source project focused

[1] Xinyu Chen, Lijun Sun. Bayesian temporal factorization for multidimensional time series prediction. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44 (9): 4659-4673. https://doi.org/10.1109/TPAMI.2021.3066551

[2] Xinyu Chen, Chengyuan Zhang, Xiaoxu Chen, Nicolas Saunier, Lijun Sun. Discovering dynamic patterns from spatiotemporal data with time-varying low-rank autoregression. IEEE Transactions on Knowledge and Data Engineering, 2023. Early access. https://doi.org/10.1109/TKDE.2023.3294440

on spatiotemporal traffic data modelling[3], demonstrating its impact with well-documented Python implementations, reproducible experiments, and state-of-the-art machine learning algorithms.

Personally, I aspire to develop my academic career with a focus on AI and its applications in transportation science, urban science, and sptiotemporal data modelling. Being a Schmidt postdoctoral fellow could be a strategic move for my career development because of the following reasons. (**Skill development**) This fellowship offers an opportunity to acquire new skills and expertise. I can enroll in machine learning courses and seminars at the University of Toronto, enhancing my understanding of AI techniques. (**Research experience**) This fellowship enables me to conduct cutting-edge research in AI, data science, urban science, and transport systems, making me a more competitive candidate in the academic job market. (**Interdisciplinary background**) By fostering collaboration across various disciplines, including computer science, statistics, mathematics, social science, and transportation engineering, this fellowship broadens my knowledge and network. Such interactions are valuable for future collaborations and career opportunities. (**Research outputs**) Successful completion of this fellowship can result in high-impact publications and research outcomes, further enriching my academic and professional portfolio. In summary, the opportunity to develop skills, expand knowledge, collaborate across disciplines, and produce impactful research outcomes through this fellowship is a crucial step in my career journey, contributing significantly to my academic career development.

Finally, I am deeply committed to advancing the principles of equity, diversity, and inclusion throughout my academic and professional journey. My experiences, both personal and academic, have reinforced my belief in the value of diverse perspectives and the importance of creating an inclusive and welcoming environment for all members of the academic community. In my role as a Schmidt postdoctoral fellow, I pledge to actively contribute to the promotion of equity, diversity, and inclusion at the University of Toronto. This commitment encompasses several key aspects. (**Mentorship & support**) I am committed to serving as a mentor for individuals from underrepresented backgrounds in my field. I will actively seek opportunities to mentor and provide guidance to students. (**Inclusive research & teaching**) I will ensure that my research and teaching practices reflect a commitment to diversity and inclusivity. This includes promoting diverse voices and perspectives in my research projects and creating an inclusive learning environment where all students feel valued. (**Collaboration**) I will actively engage in collaboration that promotes diversity and inclusion within the academic community at the University of Toronto. This includes participating in workshops and events to contribute to the ongoing dialogue on these important topics. (**Self-education**) I am dedicated to ongoing education and self-improvement to better address issues related to bias, discrimination, and systematic inequalities. I firmly believe that a diverse and inclusive academic environment not only benefits individuals from underrepresented backgrounds but also enriches the entire community by fostering creativity, innovation, and excellence. I am looking forward to being an active participant in creating such an environment at the University of Toronto and beyond.

Thank you for your consideration for my application!

Yours sincerely,

Xinyu Chen

[3]See *transdim* at `https://github.com/xinychen/transdim` on GitHub with 1,000+ stars.

Significance of Leadership Contributions

I have strong leadership experiences that are impactful and relevant to this fellowship. First, I am a staunch advocate of open-source and reproducible research. Over the past few years, I have contributed to various open-source projects using various programming languages on the GitHub platform[1]. Notably, I served as a project leader in a multi-institutional research team dedicated to addressing spatiotemporal traffic data imputation and forecasting in the transportation field. Our primary focus was on the development of machine learning algorithms (e.g., matrix/tensor completion models) for spatiotemporal traffic data, such as traffic flow data and human mobility data. In this role, I coordinated research efforts, facilitated communication among team members, and ensured that project milestones were met. Recognizing the value of open collaboration, I proposed making our project publicly available on GitHub (see `https://github.com/xinychen/transdim`). This decision facilitated collaboration—through contributions to the repository—allowed for the monitoring of research progress and enabled the broader community to follow our studies. Our research work has led to several high-impact publications, cited more than 800 times on Google Scholar. Additionally, our GitHub project has demonstrated a significant influence on shaping the direction of spatiotemporal traffic data modelling, addressing real-world challenges, and contributing to advancements in AI algorithms. With over 1,000 stars and 250 forks, our GitHub project has garnered substantial attention. Leading this project has honed my leadership, project management, and networking skills. It has also expanded my professional network and increased my visibility within my research community. This experience aligns with my career goal of becoming a research leader, where I aspire to lead multidisciplinary teams in addressing complex scientific questions.

Second, I am always eager to share mathematics and computer science-related content, providing intuitive explanations in various formats. Over the past three years, I have authored more than 30 blog posts on Medium[2], covering a wide range of topics including:

- Matrix computations (e.g., randomized singular value decomposition, matrix factorization, and tensor decomposition).
- Time series modelling (e.g., vector/matrix/tensor autoregression and dynamic mode decomposition).
- Convex/nonconvex optimization (e.g., the conjugate gradient method).
- Geospatial data analysis (e.g., climate data, weather data, energy consumption data, and urban mobility data).
- Scientific computing with NumPy/CuPy.

These blog posts have collectively garnered over 80,000 views, with two of them surpassing 10,000 views. I have received valuable feedback from readers and followers, which has helped me correct mistakes and enhance the overall quality of my blog posts. Furthermore, this experience has allowed me to establish collaborations with some people, leading to joint contributions to open-source projects. In fact, this experience has not only deepened my understanding of machine learning and data science but has also highlighted the importance of sharing ideas and knowledge. It has significantly increased my visibility within my research community, aligning perfectly with my career goal of becoming an excellent researcher.

[1] GitHub homepage: `https://github.com/xinychen`

[2] Medium homepage: `https://medium.com/@xinyu.chen`

Xinyu Chen (PhD Candidate)

Last update on October 1, 2023

Current Research Interests

- Machine Learning
- Matrix/Tensor Computations
- Low-Rank Models
- Spatiotemporal Data Modeling
- Missing Data Imputation
- Time Series Analysis
- Intelligent Transportation
- Smart Cities
- Human Mobility

Contact Information

chenxy346@gmail.com
https://xinychen.github.io (homepage)
xinychen
chenxy346
Google Scholar 879 citations (h-index: 10 & i10-index: 10)

Biography

In Fall 2023, I will finish my PhD from University of Montreal (UdeM), with support from the IVADO PhD Excellence Scholarship and the CIRRELT PhD Excellence Scholarship. My PhD research focuses on machine learning, spatiotemporal data modeling, and intelligent transportation systems.

Education

PhD in Civil Engineering (Transportation) 2020.08 – 2023.12 (expected)
Polytechnique Montreal, ***University of Montreal*** Montreal, Canada
IVADO PhD Excellence Scholarship & CIRRELT PhD Excellence Scholarship

- Thesis: *Matrix and Tensor Models for Spatiotemporal Traffic Data Imputation and Forecasting*
- Advisor: Nicolas Saunier (full professor at Polytechnique Montreal)
- Co-advisor: Lijun Sun (associate professor at McGill University)

Master's degree in Traffic Information Engineering & Control 2016.08 – 2019.06
Sun Yat-Sen University Guangzhou, China
Outstanding Thesis Award (top 2% in total)

- Thesis: *Imputing Spatiotemporal Missing Traffic Data by Bayesian Tensor Factorization Models*
- Advisor: Zhaocheng He (full professor)

Bachelor's degree in Traffic Engineering 2012.09 – 2016.06
Guangzhou University Guangzhou, China

- Thesis: *Modeling Vehicles' Time Headway with Log-Normal and Power-Law Distribution*
- Advisor: Xiaodong Zang (full professor)

Honours and Awards

- *CIRRELT PhD Excellence Scholarship* ($5,000) 2021.12
- *IVADO PhD Excellence Scholarship* ($100,000, by Institute for Data Valorisation) 2020.04
- *Outstanding Thesis Award* (by Sun Yat-Sen University) 2019.06
- *National Scholarship* (by Ministry of Education of China) 2018.11

Refereed Journal Papers

Google Scholar: https://scholar.google.com/citations?user=mCrW04wAAAAJ&hl

◆ **First-author papers** (4 papers cited above 100 times)

8. Xinyu Chen, Chengyuan Zhang, Xiaoxu Chen, Nicolas Saunier, Lijun Sun (2023). **Discovering dynamic patterns from spatiotemporal data with time-varying low-rank autoregression**. *IEEE Transactions on Knowledge and Data Engineering*. Early access.
   https://doi.org/10.1109/TKDE.2023.3294440
   JCR-Q1 IF: 8.9 top-tier

7. Xinyu Chen, Lijun Sun (2022). **Bayesian temporal factorization for multidimensional time series prediction**. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 44 (9): 4659–4673.
   https://doi.org/10.1109/TPAMI.2021.3066551
   JCR-Q1 IF: 23.6 top-tier 100+ citations
   ESI hot paper (top 0.1%) ESI highly cited paper (top 1%)

6. Xinyu Chen, Mengying Lei, Nicolas Saunier, Lijun Sun (2022). **Low-rank autoregressive tensor completion for spatiotemporal traffic data imputation**. *IEEE Transactions on Intelligent Transportation Systems*. 23 (8): 12301–12310.

https://doi.org/10.1109/TITS.2021.3113608
JCR-Q1 IF: 8.5 top-tier ESI hot paper (top 0.1%)

5. Xinyu Chen, Yixian Chen, Nicolas Saunier, Lijun Sun (2021). **Scalable low-rank tensor learning for spatiotemporal traffic data imputation**. *Transportation Research Part C: Emerging Technologies*. 129: 103226.

https://doi.org/10.1016/j.trc.2021.103226
JCR-Q1 IF: 8.3 top-tier

4. Xinyu Chen, Jinming Yang, Lijun Sun (2020). **A nonconvex low-rank tensor completion model for spatiotemporal traffic data imputation**. *Transportation Research Part C: Emerging Technologies*. 117: 102673.

https://doi.org/10.1016/j.trc.2020.102673
JCR-Q1 IF: 8.3 top-tier

3. Xinyu Chen, Zhaocheng He, Yixian Chen, Yuhuan Lu, Jiawei Wang (2019). **Missing traffic data imputation and pattern discovery with a Bayesian augmented tensor factorization model**. *Transportation Research Part C: Emerging Technologies*. 104: 66–77.

https://doi.org/10.1016/j.trc.2019.03.003
JCR-Q1 IF: 8.3 top-tier 100+ citations

2. Xinyu Chen, Zhaocheng He, Lijun Sun (2019). **A Bayesian tensor decomposition approach for spatiotemporal traffic data imputation**. *Transportation Research Part C: Emerging Technologies*. 98: 73–84.

https://doi.org/10.1016/j.trc.2018.11.003
JCR-Q1 IF: 8.3 top-tier 200+ citations ESI highly cited paper (top 1%)

1. Xinyu Chen, Zhaocheng He, Jiawei Wang (2018). **Spatial-temporal traffic speed patterns discovery and incomplete data recovery via SVD-combined tensor decomposition**. *Transportation Research Part C: Emerging Technologies*. 86: 59–77.

https://doi.org/10.1016/j.trc.2017.10.023
JCR-Q1 IF: 8.3 top-tier 100+ citations

◆ **Co-author papers**

4. Ben-Zheng Li, Xi-Le Zhao, Xiongjun Zhang, Teng-Yu Ji, Xinyu Chen, Michael K. Ng (2023). **A learnable group-tube transform induced tensor nuclear norm and its application for tensor completion**. *SIAM Journal on Imaging Sciences*. 16 (3): 1370–1397.

http://dx.doi.org/10.1137/22M1531907

3. Lijun Sun, Xinyu Chen, Zhaocheng He, Luis F. Miranda-Moreno (2021). **Routine pattern discovery and anomaly detection in individual travel behavior**. *Networks and Spatial Economics*. 35.

http://dx.doi.org/10.1007/s11067-021-09542-9

2. Pu Ren, Xinyu Chen, Lijun Sun, Hao Sun (2021). **Incremental Bayesian matrix/tensor learning for structural monitoring data imputation and response forecasting**. *Mechanical System and Signal Processing*. 158: 107734.

https://doi.org/10.1016/j.ymssp.2021.107734

1. Zhaocheng He, Kaiying Chen, Xinyu Chen (2018). **A collaborative method for route discovery using taxi drivers' experience and preferences**. *IEEE Transactions on Intelligent Transportation Systems*. 19 (8): 2505–2514.

http://doi.org/10.1109/TITS.2017.2753468

Conference Papers

WCTR 2023: Xinyu Chen, Zhanhong Cheng, Nicolas Saunier, Lijun Sun (2023). **Laplacian convolutional representation for traffic time series imputation** (presentation only). *Proceedings of the World Conference of Transport Research*.

TRB 2023: Xinyu Chen, Chengyuan Zhang, Lijun Sun, Nicolas Saunier (2023). **Nonstationary temporal matrix factorization for sparse traffic time series forecasting** (presentation only). *The 102nd Annual Meeting of Transportation Research Board*.

KDD Time Series Workshop: Xinyu Chen, Mengying Lei, Nicolas Saunier, Lijun Sun (2021). **Low-rank autoregressive tensor completion for spatiotemporal traffic data imputation** (presentation only). *The 7th SIGKDD Workshop on Mining and Learning from Time Series (MiLeTS)*.

Preprint and Submitted Papers

3. Xinyu Chen, Zhanhong Cheng, Nicolas Saunier, Lijun Sun (2022). **Laplacian convolutional representation for traffic time series imputation**. arXiv: 2212.01529.
2. Xinyu Chen, Chengyuan Zhang, Xi-Le Zhao, Nicolas Saunier, Lijun Sun (2022). **Nonstationary temporal matrix factorization for multivariate time series forecasting**. arXiv: 2203.10651.
1. Xinyu Chen, Lijun Sun (2020). **Low-rank autoregressive tensor completion for multivariate time series forecasting**. arXiv: 2006.10436.

Academic Funding

1. ***City-Scale Traffic Data Imputation and Forecasting with Tensor Learning***
   - **Authors**: Xinyu Chen, Nicolas Saunier (advisor)
   - **Link**: `https://ivado.ca/en/scholarships-and-grants/phd-excellence-scholarships/`

   IVADO PhD Excellence Scholarship $100,000   September 1, 2020

Reviewing Activities

I am serving as a reviewer for **10+** scientific journals.

- Applied Intelligence
- Big Data Research
- Expert Systems with Applications
- IEEE Intelligent Transportation Systems Magazines
- IEEE Open Journal of Signal Processing
- IEEE Sensors Journal
- IEEE Transactions on Intelligent Transportation Systems
- IEEE Transactions on Knowledge and Data Engineering
- INFORMS Journal on Computing
- Scientific Reports
- Transportmetrica B: Transport Dynamics
- Transportation Research Part B: Methodological
- Transportation Research Part C: Emerging Technologies

Professional Memberships

❒ Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (**CIRRELT**)	**Student Member**	2021 - present
❒ Institute of Electrical and Electronics Engineers (**IEEE**)	**Student Member**	2022 - present

Open-Source Projects

I am leading some innovative projects on GitHub (**3.5k+** stars & **600+** forks & **490+** followers).

◆ **Selected repositories**

transdim: Python codes for spatiotemporal data imputation and prediction using a variety of state-of-the-art machine learning (mainly including low-rank matrix and tensor methods) and deep learning models. 2018.09 - present

`xinychen/transdim`   1k+ stars

- **awesome-LaTeX-drawing**: Drawing Bayesian networks, graphical models, tensor structures, and technical frameworks in LaTeX. (Most examples are from our research papers.) — 2019.06 - present
  xinychen/awesome-latex-drawing ☆ 1.1k+ stars
- **LaTeX-cookbook**: Academic writing with LaTeX: A tutorial (in Chinese). — 2021.05 - present
  xinychen/latex-cookbook ☆ 750+ stars
- **tensor-learning**: Python codes for low-rank tensor factorization, tensor completion, and tensor regression techniques. — 2019.06 - present
  xinychen/tensor-learning ☆ 150+ stars
- **awesome-beamer**: Creating presentation slides by using Beamer in LaTeX. (Most examples are from our research.) — 2020.11 - present
  xinychen/awesome-beamer ☆ 70+ stars
- **tracebase**: Multivariate time series forecasting on high-dimensional and sparse Uber movement speed data. — 2020.11 - present
  xinychen/tracebase ☆ 40+ stars

Presentation & Talk

❒ Laplacian convolutional representation for traffic data imputation. — 2023.07

- World Conference of Transport Research (WCTR 2023)
- Montreal, Canada
- Slides: https://xinychen.github.io/slides/LCR.pdf

❒ Low-rank matrix and tensor methods for spatiotemporal traffic data modeling. — 2023.05

- Southern University of Science and Technology (SUSTech)
- Shenzhen, China
- Slides: https://xinychen.github.io/slides/traffic_data_modeling_v1.pdf

❒ Low-rank matrix and tensor methods for spatiotemporal data modeling. — 2023.04

- Sichuan University (SCU)
- University of Electronic Science and Technology of China (UESTC)
- Chengdu, China
- Slides: https://xinychen.github.io/slides/stdata_modeling.pdf

❒ Low-rank matrix and tensor factorization for speed field reconstruction. — 2023.03

- Research Group of Transport, Polytechnique Montreal
- Montreal, Canada
- Slides: https://xinychen.github.io/slides/MF_TF_SFR.pdf

❒ Spatiotemporal traffic data imputation and forecasting with tensor learning. — 2022.05

- IVADO Project Workshop
- Montreal, Canada
- Slides: https://xinychen.github.io/slides/phd_project_22summer.pdf

❒ Nonstationary temporal matrix factorization for multivariate time series forecasting. — 2022.05

- Hong Kong Machine Learning Meetup (virtual)
- Slides: https://xinychen.github.io/slides/notmf.pdf

❒ Bayesian temporal factorization for multidimensional time series prediction. — 2021.03

- IFT 6760A Course (*Matrix and tensor factorization techniques for machine learning*)
- Slides: https://doi.org/10.5281/zenodo.4693404

Skills

❒ **Language**: Chinese (native) & English (fluent)

❒ **Expertise**: Python/Matlab/Julia/R/Java; NumPy/PyTorch/CuPy; Jupyter Notebook; LaTeX; CSS/HTML.