

Chao Huang

(646)683-6745 | ch3474@columbia.edu | Homepage: henrynebula.github.io
Apt 12E, 504 W 110th Street, New York, NY, 10025

EDUCATION

Columbia University

M.S. in Data Science (GPA: 4.08/4.0)

New York, NY

Aug 2019 - Dec 2020

Relevant Coursework: Algorithms For Data Science, Statistical Inference & Modeling, Exploratory Data Analysis, Personalization Theory & Application

Tsinghua University

B.Eng. in Electronic Engineering (GPA: 3.71/4.0)

Beijing, CN

July 2019

Relevant Coursework: Fundamental Pattern Recognition, Mobile Data Mining, Theory and Methods for Statistical Inference, Operating System, Database, Computer Network Technology and Experiment, Fundamental of Digital Logic and Processor

Honor: Scholarship of Academic Excellence, Outstanding Graduate Award

SKILLS

- **Programming Languages:** Python, Java, Scala, C++
- **Machine Learning Libraries & Frameworks:** Scikit-Learn, Numpy, Scipy, Pandas, Apache Spark, PyTorch
- **Quantative Skills:** SQL, MATLAB, R

PROFESSIONAL EXPERIENCE

Platform and Content Group (PCG), Tencent

Full-time algorithm intern

Beijing, CN

June 2019 - July 2019

Content based recall algorithm for short form video recommendations

- * Applied feature engineering on video tag information along with user query keywords.
- * Trained Word2Vec embeddings of tags and reduced feature dimensions through clustering.
- * Implemented a content based recall algorithm based on video similarities, using Scala, Spark and internal machine learning platform Tesla.
- * The overall click through rate (CTR) was improved by 20% shown by online A/B test.

RESEARCH EXPERIENCE

InfoLab, University of Southern California

Undergraduate Research Assistant (Advisor: Prof. Cyrus Shahabi)

Los Angeles, CA

July 2018 - Sep 2018

Data-driven Reachability Research Using Isochrone Map

- * Visualized reachable urban areas using isochrone maps and provided useful insights for urban planning.
- * Designed a new visualization model by leveraging taxi trajectories to reflect the traffic dynamics for different time periods.
- * Created a special indexing method to support fast online query on a real world trajectory dataset with 100K+ trajectories.
- * Developed a user-friendly web UI using Node.js, Postgre and PostGIS for online visualization.

Fib Lab, Tsinghua University

Undergraduate Research Assistant (Advisor: Prof. Yong Li)

Beijing, CN

May 2017 - July 2019

Private Mobile Crowdsourcing Mechanism

- * Designed a practical private mechanism to recruit participants with maximal spatial coverage under a limited budget.
- * Utilized randomized response to protect location privacy of candidates locally with a rigorous proof of privacy guarantee.
- * Put forward a heuristic inference algorithm to maximize expected spatial coverage using obfuscated user reports.
- * Achieved at least 10% performance improvement compared with several baselines, using real-world trajectory datasets.

Private Cross-Domain Recommender System

- * Generalized the “geo-indistinguishability” obfuscation mechanism to preserve the semantic information of each Point-Of-Interest (POI).
- * Designed a confidence inference strategy to accommodate Collective Matrix Factorization (CMF) algorithm to this private scenario.
- * Evaluated our model on real-world POI datasets which are collected through Foursquare’s and Tencent’s developer API and achieved at most 140% accuracy improvement compared with single domain baselines.

PUBLICATIONS

- **Sensys’18 (Posters & Demos)** Chao Huang, Fengli Xu, Yong Li, Xinlei Chen, Pei Zhang, Poster Abstract: Locally Differentially Private Participant Recruitment for Mobile Crowdsourcing, *In the Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems*
- **IMWUT’19 (Full Paper)** Chen Gao, Chao Huang, Yue Yu, Huandong Wang, Yong Li, Depeng Jin, Privacy-preserving Cross-domain Location Recommendation, *In the Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, Volume 3 Issue 1, March 2019*
- **MDM’19 (Short Paper)** Chrysovalantis Anastasiou, Chao Huang, Seon Ho Kim, Cyrus Shahabi, Time-Dependent Reachability Analysis: A Data-Driven Approach, *2019 20th IEEE International Conference on Mobile Data Management (MDM), Hong Kong, 2019*