Harshal Chaudhari

Personal Information

Date of Birth May 14, 1991

Contact harshal@bu.edu +1-857-294-8568

Website https://www.harshalc.com

Github https://github.com/chdhr-harshal

Research Interests

Explainable A.I. Systems, Robust Optimization, Reinforcement Learning.

Education

2015–2021 **Ph.D. in Computer Science**, *Boston University*
Research Advisers: Prof. John Byers, Prof. Evimaria Terzi

2013–2015 **M.S. in Computer Science**, *Boston University*

2009–2013 **B.E.(Hons.) in Computer Science**, *Birla Institute of Technology and Science, Pilani*

Work Experience

2023–current **Senior Applied Scientist II**, *Etsy, Boston, USA*
Developed state of the art A.I. models using real time user activity to guide recommendations for users on various Etsy platforms. Involved in guiding junior scientists on developing ML models alongside interviewing and planning initiatives to power next-gen recommender systems at Etsy.

2021–2023 **Senior Applied Scientist I**, *Etsy, Boston, USA*
As a member of Recommendations ranking team, I work on developing fair, unbiased and scalable recommender system algorithms. Ranking models developed by me power recommendations across various sections of Etsy website, iOS and Android smartphone applications.

2018–2021 **Applied Scientist (part-time)**, *Zillow Group, Seattle, USA*
Research collaboration with the AI Relevance team, we explore the problem of developing a low-latency and high-quality personalized recommendation system that is robust to data imperfections while simultaneously ensuring fairness for protected groups of consumers.

2018 **Applied Scientist Intern**, *Zillow Group, Seattle, USA*
Developed an unsupervised scalable framework to identify the state of user in their home-buying journey based upon their interaction history on the real-estate marketplace Zillow. Injecting the features derived from identification of the state of journey of a user into the personalized recommendation platform at Zillow results in a significant improvement on the key metrics.

2015 **Data Science Intern**, *Amplero, Seattle, USA*
Developed a user simulator for a personalized marketing platform, Amplero, developed by Globys Inc. The characteristics of the simulated users are derived from real world telecoms usage data in a probabilistic manner. It removes operational lag associated with marketing, facilitates A/B testing for various predictive models devised by Amplero. Modular design of the user simulator further enables fine tuned differential analysis of the strategies.

2013 **Research Intern**, *Siemens Corporate Research, Bangalore, India*
Developed a clone prioritization algorithm for identification of code clones and optimal resource allocation for clone refactoring. Modeled this as a multi-constrained, multi-objective Knapsack problem and investigated various heuristics in multi-criterion branch and bound algorithms in addition to evolutionary algorithms based on Pareto optimality. We developed an Eclipse plug-in for the assessment of code duplication characteristics.

Research Experience

2023 **Towards Flexibility and Robustness of LSM Trees**, VLDB Journal: Special Issue on ML and DB
Co-authors: Andy Huynh, Prof. Evimaria Terzi, Prof. Manos Athanassoulis
Log-Structured Merge trees (LSM trees) are increasingly used as the storage engines behind several data systems, frequently deployed in the cloud. Operating in a shared infrastructure like the cloud comes with workload uncertainty due to fast evolving nature of modern applications. Systems with static tuning discount variability of workloads and provide inconsistent and suboptimal performance. Building upon our previous work, Endure, we introduce flexible compaction policy, viz., K-LSM – that allows us to express fine-grained hybrid compaction policies between the popular leveling and tiering policies. With exhaustive experimentation, we analyze performance improvements offered by K-LSM policy over state-of-the-art baselines.

2022 **Endure: A Robust Tuning Paradigm for LSM Trees Under Workload Uncertainty**, VLDB 2022
Co-authors: Andy Huynh, Prof. Evimaria Terzi, Prof. Manos Athanassoulis
Modern LSM-tree backed key-value stores co-tune merge policies, buffer sizes and the false positive rates for the Bloom filters across different levels of LSM-tree. These systems typically maximize throughput associated with updates, point and range lookup queries for fixed expected workloads. However, the analytically obtained optimal design-parameters for these systems are not always feasible. In this work, we introduce Endure – a new paradigm for tuning LSM trees in the presence of workload uncertainty. Robust tunings output by Endure lead up to a 5x improvement in throughput.

2021 **Fleet Management Strategies for Urban Mobility-on-Demand Systems**, Ph.D. Thesis
In recent years, the paradigm of personal urban mobility has radically evolved as an increasing number of Mobility-on-Demand (MoD) systems continue to revolutionize urban transportation. Hailed as the future of sustainable transportation, with significant implications on urban planning, these systems typically utilize a fleet of shared vehicles such as bikes, electric scooters, cars, etc., and provide a centralized matching platform to deliver point-to-point mobility to passengers. In this dissertation, we study MoD systems along three operational directions— (1) modeling: developing analytical models that capture the rich stochasticity of passenger demand and its impact on the fleet distribution, (2) economics: devising strategies to maximize revenue, and (3) control: developing coordination mechanisms aimed at optimizing platform throughput.

2020 **A General Framework for Fairness in Multistakeholder Recommendations**, FAccTREC, RecSys 2020
Co-authors: Sangdi Lin, Ondrej Linda
Traditionally, multistakeholder recommendations problems have been formulated as integer linear programs which compute recommendations in an offline fashion, by incorporating provider constraints. Such approaches can lead to unforeseen biases wherein certain users consistently receive low utility recommendations in order to meet the global provider coverage constraints. We propose a submodular optimization based framework incorporating seller coverage objectives alongside user objectives in a real-time personalized recommender system.

2020 **Learn to Earn: Enabling Coordination Within a Ride-Hailing Fleet**, IEEE BigData 2020
Co-authors: Prof. John Byers, Prof. Evimaria Terzi
In this work, we explore the problem of maximizing earnings of drivers employed by ride-hailing platforms like Uber, Lyft, etc. Our work confirms the idea that even in a high-dimensional and big-data domain such as ride-hailing, the inherent structure of the data can be leveraged to develop a simple, interpretable, fair and highly efficient framework that aims to achieve this goal. Furthermore, we provide evidence for model robustness and generalizability using large-scale simulations based on publicly avilable New York City taxi datasets.

2018 **Markov Chain Monitoring**, SDM 2018
Co-authors: Prof. Michael Mathioudakis, Prof. Evimaria Terzi
Given an initial distribution of items over the nodes of a Markov chain, we wish to estimate the distribution of items at subsequent times. In deriving these estimates, we issue queries to retrieve partial information on the distribution of items. For different types of queries, we design efficient algorithms for picking the right queries that make our estimates as accurate as possible.

2018 **Putting Data in Driver's Seat: Optimizing Earnings for On-Demand Ride-Hailing**, WSDM 2018
Co-authors: Prof. John Byers, Prof. Evimaria Terzi
In this study, we model the passenger seeking behavior of the Uber drivers as a controlled Markov Decision Process (MDP) over a finite horizon. The parameters of this MDP are set using Uber Rider API and publicly available New York Taxi datasets. Using this model, we devise three optimal strategies for Uber drivers and evaluate them over multiple simulations of MDP. We provide a sensitivity analysis to account for uncertainties in the MDP parameters.

2017 **Impacts of free app promotion: A case study on Amazon Appstore**, WCBA 2017, TSMO 2018
Co-authors: Prof. John Byers
In this study, we investigate the longer-term consequences of free app promotions on the performance of apps on Amazon Appstore. In particular, we quantify the causal impact of such promotions on apps' future download volumes, star ratings, and sales rank using a multi-level model. In addition, we show the presence of a cross-market spillover effect of such promotions on the performance of the same apps on Google Playstore. Our results underscore a nuanced set of trade-offs for an app developer: do the benefits of running a promotion and boosting ones' sales rank warrant the lost revenue and risk of lower user ratings in the long run?

Professional Service

PC Member WSDM '24, NeurIPS '23, SDM '23, KDD '23, WSDM '23, KDD '21, KDD '20, KDD '19, WWW '18.

Reviewer ECML-PKDD '19, TKDE '18 (Journals).
ICML '23, ICML '22, ICML'21, ICML '20, WSDM '18, ICDE '18, WWW '17 and ICWSM '17 (Conferences).

Teaching Experience

CS565 **Algorithmic Data Mining**, *Instructor: Prof. Evimaria Terzi.*

CS131 **Combinatoric Structures in Discrete Mathematics**, *Instructor: John Byers.*

CS111 **Introduction to Java Programming**, *Instructor: Prof. David Sullivan.*

Awards and Scholarships

2007–2012 National Talent Search Scholarship by the Government of India.

2005–2007 Maharashtra Talent Search Scholarship by the Government of Maharashtra state.

2005–2007 Bombay Talent Search Scholarship.

Programming Experience

Programming	Tensorflow, PyTorch, Apache Spark, Python, Scala, Java, R, C/C++, SQL, etc.
Cloud Infrastructure	GCP, AWS