Tarun Kathuria

CONTACT INFORMATION	101, Jai Jawan Colony-3,	E-mail: tarunkathuria@gmail.com
	J.L.N. Marg,	Homepage: `research.microsoft.com/∼t-takat`
	Jaipur - 302018, India	Mobile: +91-8879346633

RESEARCH INTERESTS   Databases, Machine Learning and High-dimensional Statistics, Approximation Algorithms

EMPLOYMENT   **Microsoft Research India**, Bangalore   *July 2015 - Present*

Research Fellow, Algorithms and Theory Group   *Mentor: Dr. Amit Deshpande*

EDUCATION   **Indian Institute of Technology - Bombay**, Mumbai, India   *July 2011 - June 2015*

Bachelor of Technology (Honors)

- Major: Computer Science & Engineering
- Minor: Applied Statistics and Informatics
- CGPA: 9.08/10.00
- Ranked 10th in the department (among 96 students)

PUBLICATIONS   Tarun Kathuria, S. Sudarshan. "Greedy Awakens : Efficient and Provable Multi-Query Optimization". *Under review at* **PODS 2016**. Manuscript available at `http://arxiv.org/abs/1512.02568`

RESEARCH EXPERIENCE

Undergraduate Dissertation, IIT Bombay   *July 2014 - May 2015*

Efficient and Provable Algorithms for Multi-Query Optimization   Prof. S. Sudarshan

- Work in the area of multi-query optimization (finding a set of common sub-expressions to materialize in order to find the optimal plan for a batch of queries) relies on *development of heuristics* which work well in practice. *No theoretical guarantees* on the quality of solution obtained by any heuristics exist so far
- Under assumptions of **submodularity** of a reformulation of the problem which is known to work well in practice, proposed an algorithm for **unconstrained normalized, submodular maximization** when the values may be negative
- **Proved an approximation factor** for the proposed algorithm. Further showed that it is NP-hard to achieve a **better approximation ratio**. Also proposed optimizations and pruning techniques for the algorithm which preserve the theoretical guarantee [**Paper submitted to PODS 2016**]
- Integrating the above into **PyroJ**, a *Volcano/Cascades*-based query optimizer developed at **IIT Bombay**.

Research Project, Microsoft Research India   *July 2015 - November 2015*

Evaluation of Anomaly Detectors in Data Analytics Platforms   Dr. S. Sellamanickam

- Anomaly detectors for many modern data analysis platforms are created primarily using *loose domain knowledge*. Many such detectors may not be useful and may be misleading.
- It is, thus, desirable to *rank the detectors based on quality* and prune out bad detectors, in the **absence of accurately labelled data** from past anomalies
- Investigated properties of a restricted set of anomalies and devised an *ensemble-based model for ranking* as well as creating a more accurate detector

Research Project, Microsoft Research India   *December 2015 - Present*

Classification under Low Rank and Missing Data   Dr. Amit Deshpande & Dr. Prateek Jain

- Often, for classification tasks, the training data may contain missing attribute values. Current methods make *low rank* as well as *fairly strong assumptions on the distribution* of the data
- Currently exploring approaches to solve such classifications tasks under *less restrictive assumptions*

Internship, IBM Research Labs, Bangalore   *May 2014 - July 2014*

Large Scale Topical Analysis using the Social Network   Dr. Indrajit Bhattacharya

- Proposed extensions to existing **non-parametric topic modelling algorithms** to account for various relations amongst users like geographies, friend circles, achieving better accuracy than existing models
- Devised **influence-aware topic models** to capture the effect of *user information cascades* by incorporating the **Independent Cascade Model** in these models
- Implemented a preliminary version of the above in the **Hadoop** framework for large-scale Twitter analysis

Research Project, IIT Bombay — *Autumn 2014*
Graph Learning using Orthonormal Representation — Prof. Saketha Nath & Prof. Chiranjib Bhattacharyya

- **Proved an equivalence** between *margin complexity* of one-class SVMs and *Lovasz ϑ function* on graphs
- Explored approaches to proving the **statistical consistency** of transductive learning on graphs using *orthonormal representations* using *VC-dimension* and *Rademacher complexity* based risk bounds

<div></div>

ACADEMIC
DISTINCTIONS
AND AWARDS

Awarded **AP grade** for exceptional performance in *Introduction to Numerical Analysis, Electricity & Magnetism and Introduction to Linear Algebra*; awarded to **top 2** % students at IIT Bombay

Awarded the **Institute Academic Prize** in 2012 for academic excellence at IIT Bombay

Secured 10.0/10.0 GPA in the 7th semester (Fall 2014)

Amongst the top **300 (0.1%)** to be selected for *Indian National Chemistry Olympiad, 2011*

Awarded **CBSE Merit Certificate** in *Mathematics (100% marks obtained)* & in *Computer Science* for being in the **top 0.1%** of the students taking the *CBSE 2011* high school examinations

Offered a 100 % scholarship to pursue undergraduate studies at the **Hong Kong University (HKUST)**

Awarded **A* grade** in *Mathematics, Physics, Chemistry and Information Technology (Computer Science)* in the **IGCSE 2009 (University of Cambridge International Examinations)** in the 10th grade

<div></div>

KEY ACADEMIC
PROJECTS

Subquery optimization in PyroJ
Guide: Prof. S. Sudarshan — *Autumn 2014*

- Implemented algorithms for elimination of subqueries in the **PyroJ** query optimizer at IIT Bombay using ideas like the *Apply* operator from **Execution Strategies for SQL Subqueries, Elhemali et al., SIGMOD '07**
- Implemented subquery optimization using *Magic sets* as well as by re-ordering and efficient evaluation of semi-joins and anti-joins

Twitter Sentiment Analysis
Guide: Prof. Pushpak Bhattacharya — *Spring 2014*

- Implemented a feed-forward with back-propogation neural network and applied it for sentiment analysis of about 50k Twitter annotated tweets
- Implemented various syntactic, semantic and stylistic features for better feature selection. Achieved the highest accuracy in the class for the same

Chess Titans
Guide: Prof. Amitabha Sanyal — *Spring 2013*

- Developed a one player chess game, in PLT Scheme using the DrRacket in-built GUI Toolkit
- Implemented the **Minimax algorithm** with $\alpha - \beta$ **pruning** and various clever heuristics for move determination

Detecting and Classifying Geometric Shapes
Guide : Prof. Varsha Apte — *Autumn 2012*

- Developed an application to detect various shapes like *conics, lines, polygons* in a given image
- Implemented famous techniques like **Hough Transforms, Bresenham Line Drawing and Harris Corner Detection** along with novel approaches based on regression analysis and heuristics

Estimation of Gamma Parameters with Censored Samples
Guide: Prof. Siuli Mukhopadhyay — *Autumn 2013*

- Studied methods for estimation of gamma distribution parameters in censored data
- Implemented line search optimization techniques and Newton's numerical approximation methods to efficiently find the point and interval **Maximum Likelihood** estimates in Matlab

On the Complexity of Linear Prediction — *Autumn 2014*

Surveyed and presented literature on *risk, margin and covering number bounds* of linear classification algorithms using **Rademacher** & **Gaussian complexities** as part of an *Advanced Machine Learning* course

Computational Humor — *Autumn 2014*

Surveyed recent literature on **Computational Humor Recognition and Generation** and presented the same to a class of 90 students as part of an *AI* course, earning the **highest mark** for the same

TEACHING EXPERIENCE	**Undergraduate Teaching Assistant**	*Summer 2013, Spring 2014, Spring 2015*

Undergraduate Teaching Assistant — *Summer 2013, Spring 2014, Spring 2015*
Course : Introduction to Numerical Analysis — *Prof. S. Baskar*
Assisted the professor in *setting question papers and model solutions* for examinations, conducting *problem solving* sessions and invigilating for examinations for **three offerings** of the course

Undergraduate Teaching Assistant — *Spring 2013*
Course : Electricity & Magnetism — *Prof. Tomy C.V.*
Assisted the professor in *setting question papers and model solutions* for examinations, conducting *problem solving* sessions and invigilating for examinations of the course

High School Teaching Volunteer — *Winter 2013*
Course : Basic Programming in Java
Taught a class of 30 students in high school about fundamentals of *Java programming* and *algorithmic thinking*

TECHNICAL SKILLS

Programming	C++, Java, Python, Julia, Ruby, Haskell, Scheme (Lisp)
Databases	PostgreSQL, Hive, Hyracks, HBase, Neo4j
Web Development	JavaScript, Rails, Django
Software Packages	Hadoop, Mahout, Matlab, OpenCV

RELEVANT COURSES UNDERTAKEN

Core: Statistical Techniques in Data Mining, Implementation of Relational Database Systems, Advanced Databases, Foundations of Machine Learning, Topics in Machine Learning, Markov Decision Processes, Game Theory, Artificial Intelligence, Linear Optimization, Convex Optimization, Operating Systems, Networks, Computer Architecture, Automata Theory, Data Structures, and Algorithms, Algorithm Design

Breadth: Calculus, Linear Algebra, Differential Equations, Numerical Analysis, Electricity and Magnetism, Chemistry, Psychology

INDEPENDENT STUDY

I have studied the following courses on coursera.org

- Machine Learning
- Probabilistic Graphical Models
- Algorithms - I & II
- Mining Massive Datasets

EXTRA CURRICULAR ACTIVITIES

- Participated in **Mozilla's MozBoot 2014**, an overnight code contribution sprint to bootstrap developer contribution to Mozilla's large open source projects like *Firefox* & *Servo*
- Completed a 1 year **Guitar** course offered by *National Sports Organization* at IIT Bombay
- Solved Rubik's Cube as part of IIT Bombay's successful Guinness World Record attempt for maximum number of people simultaneously solving the Rubik's Cube
- Served as Editor of my high-school magazine
- Won various national debate competitions at the high school level
- Playing Chess and solving game puzzles like Sudoku & Kakuro

REFEREES

Professor S. Sudarshan
Head of Dept. (CSE)
IIT Bombay
Mumbai, India
e-mail: sudarsha@cse.iitb.ac.in

Dr. Indrajit Bhattacharya
Senior Researcher
IBM Research Labs
Bangalore, India
e-mail: indrajitb@in.ibm.com

Professor S. Baskar
Assistant Professor (Math)
IIT Bombay
Mumbai, India
e-mail: baskar@math.iitb.ac.in

Professor Sachin Patkar
Professor (EE)
IIT Bombay
Mumbai, India
e-mail: patkar@ee.iitb.ac.in