IUPUI

SCHOOL OF SCIENCE

A PURDUE UNIVERSITY SCHOOL
Indianapolis

December 14, 2018

RE: Approval of the MS in Computational Data Science

The IUPUI School of Science has developed an innovative new master's program in Computational Data Science to educate graduate students on the Indiana University Purdue University Indianapolis campus. The program proposal was developed and approved by the faculty in the Computer and Information Sciences Department as well as the Dean for the School of Science. The Indiana University Purdue University Indianapolis Graduate Affairs Committee reviewed and approved the new MS program along with the campus administration in Indianapolis. The proposal was reviewed and approved by faculty and staff in the Graduate School at Purdue University. The complete proposal for this new MS degree has been successfully routed and approved using the Purdue Curriculog system. This memo confirms the approval of this new degree proposal.

Shiaofen Fang, Ph.D.
Department Chair, Computer and Information Sciences
School of Science
Indiana University Purdue University Indianapolis

Simon Rhodes, Ph.D.
Dean
School of Science
Indiana University Purdue University Indianapolis

Janice Blum, Ph.D.
Associate Vice Chancellor for Graduate Education
Indiana University Purdue University Indianapolis
Associate Dean
Purdue University Graduate School

Linda Mason, Ph.D.
Dean
Purdue University Graduate School

Jay T. Akridge
Provost and Executive Vice President for Academic Affairs and Diversity
Purdue University

Mitchell E. Daniels, Jr.
President
Purdue University

INTRA-AGENCY ADVISORY AND DELIBERATIVE MATERIAL

MEMORANDUM

Executive Summary of Upcoming Board Review or Action Item

DATE: December 14, 2018
TO: Board of Trustees
FROM: Dr. Mihran Tuceryan, Primary Contact, (317) 274-9736; tuceryan@iu.edu
CC: Nicole Wittlief, Secondary Contact, (317) 274-3883; wittlief@iupui.edu
SUBJECT: Proposal for M.S. in Computational Data Science, IUPUI Computer Science

Purpose:

□ This item is recommended for discussion by the Board in executive session.
□ This item is recommended for discussion by the _____ Committee at its meeting.
X This item will require a formal vote by the Academic & Student Affaurs Committee at its meeting.
□ This item will require a formal vote by the Board in the Stated Meeting.
□ This item will be presented in a formal resolution for action at the Stated Meeting.

Attachments: The following attachments accompany this memo: No attachments

Executive Summary: The Departments of Computer & Information Science and Mathematical Sciences of the IUPUI School of Science propose a Master of Science program in Computational Data Science. This program will utilize selected existing graduate courses offered by the two Departments, with related, complimentary topics in Computer Science and Statistics, in a sequence that builds to yield deeper understanding of different theoretical and applied platforms of data science.

The objective of this program is to prepare students to enter the workforce in the rapidly advancing field of data science, an interdisciplinary domain that cuts across computer science and statistics, by providing a solid, comprehensive background in the related topics of theory and their applications. Data science is often seen as a very applied topic, and students are trained on specific computational tools for domain-driven tasks. On the contrary, the goal of this program is to provide all skills necessary that will enable students to be flexible and competitive in today's job market by gaining deep understanding of theory, implementation (e.g., algorithms and appropriate computing languages), as well as the inherent "nature" of different data modalities, such as classification and prediction challenges on specific data (e.g., sparse and/or incomplete data).

Data science is broadly defined as a cross-disciplinary field, on the border between Computer Science and Statistics, and involves data-driven knowledge discovery in terms of pattern analysis and prediction. Being an applied science field, data science includes not only the foundational topics from Statistics, such as regression and sequence analysis, Machine Learning, and Data Mining, but also related topics targeting the applied component of data analysis, such as database systems, information visualization, and "big data" analytics. What completes data science training is the computational/algorithmic thinking, which primarily involves the implementation aspect of methodologies, namely the ability to translate the mathematical description of a method into computational terms, to design the appropriate algorithms, and to study/improve implementation performance in terms of accuracy and computational efficiency (time). Given the above aspects of data science, one can understand that training students in this field can be challenging and it requires a careful choice of Computer Science and Statistics courses that not only complement each other, but are also in a sequence that incrementally yields unique expertise.

Executive Summary

The Departments of Computer & Information Science and Mathematical Sciences of the IUPUI School of Science propose a Master of Science program in Computational Data Science. This program will utilize selected existing graduate courses offered by the two Departments, with related, complimentary topics in Computer Science and Statistics, in a sequence that builds to yield deeper understanding of different theoretical and applied platforms of data science.

The objective of this program is to prepare students to enter the workforce in the rapidly advancing field of data science, an interdisciplinary domain that cuts across computer science and statistics, by providing a solid, comprehensive background in the related topics of theory and their applications. Data science is often seen as a very applied topic, and students are trained on specific computational tools for domain-driven tasks. On the contrary, the goal of this program is to provide all skills necessary that will enable students to be flexible and competitive in today's job market by gaining deep understanding of theory, implementation (e.g., algorithms and appropriate computing languages), as well as the inherent "nature" of different data modalities, such as classification and prediction challenges on specific data (e.g., sparse and/or incomplete data).

Data science is broadly defined as a cross-disciplinary field, on the border between Computer Science and Statistics, and involves data-driven knowledge discovery in terms of pattern analysis and prediction. Being an applied science field, data science includes not only the foundational topics from Statistics, such as regression and sequence analysis, Machine Learning, and Data Mining, but also related topics targeting the applied component of data analysis, such as database systems, information visualization, and "big data" analytics. What completes data science training is the computational/algorithmic thinking, which primarily involves the implementation aspect of methodologies, namely the ability to translate the mathematical description of a method into computational terms, to design the appropriate algorithms, and to study/improve implementation performance in terms of accuracy and computational efficiency (time). Given the above aspects of data science, one can understand that training students in this field can be challenging and it requires a careful choice of Computer Science and Statistics courses that not only complement each other, but are also in a sequence that incrementally yields unique expertise.

Section I—ICHE proposal

Cover Page

INSTITUTION: Indiana University-Purdue University Indianapolis (IUPUI)

CAMPUS: Indianapolis

COLLEGE: Purdue School of Science

DEPARTMENT/SCHOOL: Computer and Information Science, Mathematical Sciences

DEGREE PROGRAM TITLE: Master of Science (M.S.) in Computational Data Science

SUGGESTED CIP CODE: 30.3001

PROJECTED DATE OF IMPLEMENTATION: August 2018

1) Characteristics of the Program

a) Campus(es) Offering Program: Indiana University Purdue University Indianapolis (IUPUI)

b) Scope of Delivery Specific Sites or Statewide): IUPUI

c) Mode of Delivery (Classroom Blended, or Online): Classroom

d) Other Delivery Aspects (Co-ops, Internships, Clinicals, Practics, etc.): Students will be required to complete a Capstone course.

e) Academic Unit Offering Program: School of Science, IUPUI

2) Rationale for Program

a) Institutional Rationale (e.g. Alignment with Institutional Mission and Strengths)

Indiana University-Purdue University Indianapolis (IUPUI)'s mission includes education, research, and service to community. As part of this mission, it needs to educate students with skills and knowledge that aligns with existing or emerging job markets. Data science is one such emerging knowledge area and job market. Therefore, the MS in Computational Data Science degree aligns well with this mission, as Data Science has been an emerging discipline in recent years worldwide, nationwide, and statewide in Indiana.
Offering this degree at IUPUI School of Science will complete the tri-city area (Purdue University West Lafayette, Indiana University Bloomington, and IUPUI at Indianapolis) knowledge infrastructure feeding skilled workers in the Central Indiana area.

b) State Rationale

Data science has been developing for a while in certain high tech geographical areas such as Silicon Valley. However, Indiana and, in particular, Indianapolis area has seen the growth of data science and data and business analytics related companies in recent years. In addition to the existing big health related companies such as Lilly and Anthem, there have been more technological oriented companies expanding in the central Indiana region (e.g., Interactive Intelligence, Salesforce, etc.) and some start ups such as High Alpha. All of these companies will need to hire workers with the knowledge intensive skill sets.
The two Purdue Departments, Computer & Information Science and Mathematical Sciences, have the mission to educate students at the undergraduate and graduate levels that will meet the needs of local industries and businesses. The newly proposed MS in Computational Data Science program is addressing this need.

c) Evidence of Labor Market Need

i) National, State, or Regional Need

Around the nation and internationally, there are relatively few data science programs, and even fewer that incorporate training, on-site or online, in all above aspects. In Indiana, IU Bloomington has an established MS degree in data science, on-campus and online, while in Indianapolis there is no MS degree in data science offered. Also, the unique profile of the

IUPUI campus, its location, and the industry job opportunities in the broader Indianapolis area give this program the highest potential of success in training Indiana's workforce and beyond.

In terms of local jobs, a LinkedIn search of Indiana jobs for analytics or data science currently returned over 1000 results. Examples included:

- Analyst-Data Science at Indiana University Health
- Data Scientist - Applied Technology Crowe Horwath
- Senior Data Scientist Viral Launch
- Advanced Analytics Data Scientist Dow Chemical Company
- Data Scientist State of Indiana
- Data Scientist MTP Client
- Wind Energy Data Scientist Sentient Science
- Statistician - Business Analytics Eli Lilly

The IUPUI MS in data science will focus recruitment and job placement efforts in the greater Indianapolis area to help fulfill the data science workforce needs of the area. We have received letters of support from local businesses, as well as appropriate departments at Purdue West Lafayette and Indiana University Bloomington, that provide additional context to the data science need in Indiana and the advantage of having a local program in Indianapolis. We summarize the letters in section vi) below.

ii) Preparation for Graduate Programs or Other Benefits

This program aims at providing the best possible trade-off between rigorous training, duration of studies, and degree quality. The goal is to create a new brand of professionals that can identify data-specific problems, provide solutions in a principle way, using established or new methods, be proficient in computational implementation, while still having a global view of the domain-specific objective and the application constraints, limitations, and/or requirements.

This program also aims at training students that can be flexible in the job market of the future, by providing the essential fundamental background (e.g., in Machine Learning) for upcoming technological trends. For example, in the article "Will we soon no longer need data scientists?" by B. Marr, published in *Data Science Central* (May 6, 2016), the question is raised: whether the skill sets provided by current data science programs are sufficient for following the technological advances. For instance, it is mentioned that "A traditional data scientist might receive training in R or SAS or whatever tool their school uses, but we found in the 'citizen analyst' area, they were often being given the wrong tools where they were required to guess the right answer, and then test their guess." The basis of this concern comes from the tool and/or application-specific training many students receive, while lacking the background knowledge in core topics, such as Machine Learning, algorithms and data structures, as well as fundamental theories in Statistics. This article supports the idea that "cognitive technologies" will soon render traditional data scientists, as we know them today, practically of little or no need, as the machine will be capable of making choices of what methods should be used and even program itself to implement/program them. Therefore, the description of the data scientist should go beyond the typical analysis of data, and will soon focus on tasks related to how to make a machine make the "right choices". While fulfilling the current industry needs, this program also intends to provide sufficient knowledge for keeping up with such trends.

iii) Summary of Indiana Department of Workforce Development and/or U.S. Department of Labor Data

According to Bureau of Labor Statistics, Computer and Information Scientist jobs are projected to grow 19% from 2016 to 2026[1]. More specifically, "workers who use big data are employed by many kinds of institutions and in many different industries: government, businesses, financial institutions, healthcare, scientific research facilities, colleges and universities, and others. The collection and use of big data continues to expand in all of these."[2] All of these institutions are represented in the Central Indiana and Indianapolis area and, therefore, the data science jobs are also expected to grow accordingly. Similar projections are made at the local level in Indiana: 10-year growth in Computer and Mathematical Occupations is projected to be 18.3%[3].

iv) National, State, or Regional Studies

Data science professionals are currently in high demand in a wide range of industries, spanning from tech companies to pharmaceutical, biomedical, automotive, financial, etc. Any translation of the job market in the broader field of data analysis essentially leads to the conclusion that training in data science is and will be, for the next several years, among the most desired technical skills.

An IBM jobs report[4] estimates that 2.7 million data science positions will be created by 2020. Machine learning and data scientist are the top 2 "emerging jobs" in a LinkedIn U.S. jobs report[5]. "Businesses will need one million data scientists by 2018"; "Universities can't crank out data scientists fast enough to keep up with business demands"; Deepening shortage of data science talent are trends shaping business in 2016"; "Data scientists have been called "unicorns" because finding the right person with the right set of skills – including coding, statistics, machine learning, database management, visualization techniques, and industry-specific knowledge – could be practically impossible". These are indicative quotes that a potential student finds on the web before deciding about attending a data science program; they reveal what most people in academia and industry already know: "making sense of data", or formally data-driven knowledge discovery, as encoded by the title and essence of data science, is a very popular field of study that can yield high return in one's career. A. Coggine, guest author in *techzone360.com* wrote the article "Top most in-demand, highest paying tech jobs in 2017", where data scientists are at the top of the list, earning more than $100K annually. A follow-up articles in *monster.com*, by K. Felicetti, justified Coggine's conclusions with "data is bigger than ever" and "code is still king". Data scientist jobs in specific sub-fields, such as data architecture, are shown to also yield annual incomes more than $100K (salary.com: data architect, median annual income: $100,717). Indiana and the broader Indianapolis are no exceptions to this national trend.

v) Surveys of Employers or Students and Analyses of Job Postings

A simple search in online public domains reveals the high demand of skilled Computer Scientists; for instance, in November 2017 a nation-wide search in indeed.com yielded 190,444

---

1 https://www.bls.gov/ooh/computer-and-information-technology/computer-and-information-researchscientists.htm#tab-6

2 https://www.bls.gov/careeroutlook/2013/fall/art01.pdf

3 http://www.hoosierdata.in.gov/FD/overview.aspx

4 https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=IML14576USEN

5 https://economicgraph.linkedin.com/research/LinkedIns-2017-US-Emerging-Jobs-Report

Computer Science jobs (1,819 in Indiana and 873 in Indianapolis). Nationwide there are 23,123 data scientist jobs (in Indiana there are 130 data science jobs and in Indianapolis 72). Of those, 14,211 are machine learning jobs and of those 3198 are entry level. In the same month, there were thousands of positions posted nationwide for related fields, such as Computational Science (1,485 jobs), Computational Statistics (1141 jobs), and Data Analytics (60,079 jobs).

vi) Letters of support

Four letters of support are included from IUPUI in the Appendix—one from Dr. Mathew J. Palakal, Executive Associate Dean of the IU School of Informatics and Computing at IUPUI, one from Dr. Feng Li, Interim Chair and Associate Professor in the Department of Computer Information and Graphics Technology at IUPUI, one from Dr. Paul K. Halverson, Founding Dean and Professor of the IU Richard M. Fairbanks School of Public Health and the fourth from Dr. Brian King, Chair and Associate Professor in the Department of Electrical and Computer Engineering at IUPUI.

We have also received letters of support from local businesses, as well as appropriate departments at Purdue West Lafayette and Indiana University Bloomington, that provide additional context to the data science need in Indiana and the advantage of having a local program in Indianapolis. Tyler Foxworthy, Chief Scientist at DemandJump, states that he has "been generally disappointed with the availability of qualified early career data scientists in the Midwest" and that he sees the MS at IUPUI "as a great benefit to not only my business, but to the greater tech community within the State of Indiana." The Chair of the CS Department at IU Bloomington, Amr Sabry, writes "there is a clear demand for Data Science skills in the state of Indiana. While our Bloomington school is offering both online and residential programs in Data Science, a complementary residential offering in the Indianapolis area is well-suited to the residential demand in that area."

3) Cost of and Support for Program

a) Costs

i) *Faculty and Staff*: The CIS Department currently has 16 tenured or tenure-track faculty members, who already serve as graduate Advisors and Mentors. There are plans to hire 1-2 additional tenure-track faculty members in the coming years.

ii) *Facilities*: As this program is built largely from courses that already exist at IUPUI, existing learning and teaching facilities will be used to support the program.

iii) *Other Capital Costs (e.g. Equipment)*: As this program is built largely from courses that already exist at IUPUI, no additional costs are anticipated for capital equipment. Existing capital resources already in use will continue to be used.

b) Support

i) *Nature of Support (New, Existing, or Reallocated)*: As this program consists largely of courses already existing at IUPUI, mechanisms for support are already in place and currently utilized.

ii) *Special Fees above Baseline Tuition*: Students in this program will be assessed the same tuition and fees as other School of Science graduate students at IUPUI. This is expected to continue and no additional or special fees are anticipated.

4) Similar and Related Programs

a) List of Programs and Degrees Conferred

i) Similar Programs at Other Institutions

M.S. in Data Science, IU Bloomington, School of Informatics, Computing, and Engineering

ii) Related Programs at the Proposing Institution
M.S. in Applied Data Science in the School of Informatics at IUPUI. This program focuses on applications of the data science and provides students with core competencies in methods of data management, analysis, and infrastructure and high-throughput data storage. It is different than the proposed M.S. in Computational Data Science program, which will focus on the theoretical and computational foundations of Data Science.

b) List of Similar Programs Outside Indiana
M.S. in Analytics, University of San Francisco, College of Arts and Science
M.S. in Statistics: Data Science, Stanford University, Department of Statistics

c) Articulation of Associate/Baccalaureate Program Not Applicable.

d) Collaboration with Similar or Related Programs on Other Campuses Not Applicable.

5) Quality and Other Aspects of the Program

a) Credit Hours Required/Time to Completion
The curriculum requires 30 credits in total that can be completed in three semesters. There are 9 credits for core courses in Computer Science, 6 credits for Statistics core courses, 12 credits for elective courses from Computer Science and/or Statistics, and 3 credits for a capstone course. The students must choose at least two electives from Computer Science and at least two electives from Statistics. All courses are 3-credit courses. (See Appendix for course descriptions.)

See Section II for a sample plan of study.

b) Exceeding the Standard Expectation of Credit Hours
Not applicable.

c) Program Competencies of Learning Outcomes
The following Student Learning Outcomes describe what students should be able to do upon completion of the degree:

1) Synthesize data analysis principles across the statistical and computer sciences in topics such as pattern analysis, prediction, and big data processing. [*PGL: Demonstrate the knowledge and skills needed to participate in disciplinary studies or to enter a program to earn a more advanced degree*]
2) Construct data science algorithms, including derivation and programing implementation in a variety of languages and platforms (C++, Python, Java, SAS, R, Matlab). [*PGL: Demonstrate the knowledge and skills needed to participate in disciplinary studies or to enter a program to earn a more advanced degree*]
3) Be able to assess new programing language trends in industry, by gaining solid background in computing and algorithmic thinking. [*PGL: Think critically and creatively to evaluate literature in their field of study*]
4) Differentiate the processes from “raw data to outcome”, which spans from considering the domain-specific constraints and data characteristics (e.g., static vs sequence, sparsity, dimensionality, etc.) to efficient method implementation, as software, with desired

specifications. [*PGL: Think critically and creatively to evaluate literature in their field of study*]

5) Integrate advanced knowledge in a broad range of related topics, such as survival analysis and experimental design in Statistics, as well as databases, visualization, and data structures in Computer Science. [*PGL: Demonstrate the knowledge and skills needed to participate in disciplinary studies or to enter a program to earn a more advanced degree*]
6) Assess different solutions to specific data-specific problems. *[PGL I: Think critically and creatively to evaluate literature in their field of study; PGL II: Communicate effectively information from their field of study*]
7) Summarize state of the art data science methods and applications in scientific project reports and software documentation [*PGL: Communicate effectively information from their field of study*]

d) Assessment

The proposed program will be evaluated using the following parameters:

1. Number of applicants to the program and admitted students
2. Number of students attending
3. Academic profiles of attendees (GPAs, GRE scores, graduate degrees, previous institutions attended, ranking in previous institution)
4. Student performance in course work
5. If applicable, student research productivity (publications, presentations, etc. that may result from their capstone/term project)
6. Awards and other special recognition
7. Time to degree completion
8. Number of graduates from the program
9. Student placement: Number placed and quality of placements

Monitoring the above parameters will be the responsibilities of the Department Chair of the respective departments, the Director of Graduate Studies and the Departmental Graduate Committee. Each year, in early fall, the recruiting data (Items 1 and 2) from the previous year will be collected and summarized. Historical data (Item 3) on matriculating students will also be compiled at that time. Each annual cohort will be followed through graduation to collect the information for items 7 and 8. Course work performance will be monitored semester by semester (Item 4). To permit a long-term view of career trajectories, data regarding student research productivity, awards and job placement will be followed for three to five years post-graduation (Items 5, 6 and 9).

e) Licensure and Certification

The M.S. program in Computational Data Science is not intended to prepare students for any specific licensures or certifications.

f) Placement of Graduates

Students will have access to various career assistance resources on campus, including the Science PREPS office and many job and internship fairs conducted throughout the year. Students will also benefit from an access to Purdue University's CCO job banks and resources, including Science/Technology focused job fairs held on the West Lafayette campus.

g) Accreditation

The proposed program will not be accredited as no such mechanism exists for U.S. graduate programs in Computer Science.

6) Projected Headcount and FTE Enrollment and Degrees Conferred

NEW ACADEMIC DEGREE PROGRAM PROPOSAL SUMMARY

Institution/Location: IUPUI
Program: Master of Science in Computational Data Science
Proposed CIP Code:
Base Budget Year: 2017-18

	Year 1 2018-19	Year 2 2019-20	Year 3 2020-21	Year 4 2021-22	Year 5 2022-23
Enrollment Projections (Headcount)					
Full-time Students	6	6	6	6	6
Part-time Students	6	12	12	12	12
	12	18	18	18	18
Enrollment Projections (FTE)*					
Full-time Students	8	8	8	8	8
Part-time Students	4	8	8	8	8
	11	15	15	15	15
*Sum of rounded detail may not equal rounded totals.					
Degree Completion Projection	6	12	12	12	12

Section II—Academic Program Description (Appendix A & B format) Information reported in Section I may be referenced.

1) Curriculum and Requirements

a) Admissions Requirements

Qualified graduate students are those that meet the MS admission criteria of the Computer and Information Science Department or Mathematical Sciences Department, as implemented by the IUPUI Graduate Office: the student must have a 4-year Bachelor's degree in Computer Science, Mathematics, Data Science, Statistics, Engineering, or a related field. For those candidates that do not have background in the aforementioned fields, case-by case application evaluation will be followed. Students will be admitted to the Computer and Information Science Department if they applied to this department and will be admitted to the Mathematical Sciences Department if they have applied to the Mathematical Sciences Department; the Graduate Committee of each department will be responsible for evaluating the case-by-case applications.

Prerequisite coursework and/or degrees
4-year Bachelor's degree in Computer Science, Engineering, Mathematics, Statistics, or related fields. 4-year Bachelor's degree in any other area of study will be considered on a case-by-case basis, based on the coursework and corresponding grades in the applicant's transcripts, as well as on the overall potential of successfully completing this program.

GPA: Entering students are expected to have a minimum cumulative grade point average (GPA) equivalent to at least 3.00 (A=4, B=3, C=2, D=1, F=0), or equivalent.

GRE: Scores on the Graduate Record Exam (GRE) must be submitted for admission consideration.

English Proficiency Requirements: All applicants whose native language is not English are required to submit scores for TOEFL or IELTS. An overall TOEFL IBT score of 80 or higher, or an IELTS band score of 6.5 or higher is required. Applicants submitting TOEFL scores must also meet the following section minimum requirements in addition to the minimum Total requirement: 18 Writing, 18 Speaking, 14 Listening, 19 Reading.

b) Curriculum Requirements

The proposed curriculum requires 30 credits in total that can be completed in three semesters. There are 9 credits for core courses in Computer Science, 6 credits for Statistics core courses, 12 credits for elective courses from Computer Science and/or Statistics, and 3 credits for capstone course. The students must choose at least two electives from Computer Science and at least two electives from Statistics. All courses are 3-credit courses. (See below for course descriptions.)

Successful completion of the program requires a minimum plan of study GPA of 3.0, the minimum grade in any course is C, and the maximum number of courses with grades of C or C+ is two.

Core courses:

Dept.	Course Title		Credits
CSCI	59000	Introduction to Data Science	3
CSCI	57300	Data Mining	3
CSCI	57800	Statistical Machine Learning	3
STAT	51200	Applied Regression Analysis	3
STAT	52900	Applied Decision Theory and Bayesian Analysis	3

Capstone courses:

CSCI	69500	MS Capstone Project	3
STAT	59800	Topics in Statistical Methods	3

Elective courses:

Dept.	**Course Title**		**Credits**
CSCI	52000	Computational Methods in Analysis	3
CSCI	54100	Database Systems	3
CSCI	55200	Advanced Graphics & Visualization	3
CSCI	58000	Algorithm Design, Analysis, and Implementation	3
CSCI	59000	Large-scale Machine Learning	3
CSCI	59000	High Performance Computing	3
STAT	51400	Design of Experiments	3
STAT	52000	Time Series and Applications	3
STAT	52300	Categorical Data Analysis	3
STAT	52400	Applied Multivariate Analysis	3
STAT	52501	Generalized Linear Models	3
STAT	53600	Introduction to Survival Analysis	3

Students can choose one of the two capstone courses, from Computer Science (CSCI 69500) or Statistics (STAT 59800), depending on the topic and/or the advisor of their choice.

The course sequence is crucial for successfully completing this program. While the elective courses can be taken at any time during the program period, the core courses are to be offered in the specified sequence:
(1) CSCI 59000: Introduction to Data Science, STAT 51200: Applied Regression Analysis
(2) CSCI 57300: Data Mining, STAT 52900: Applied Decision Theory and Bayesian Analysis
(3) CSCI 57800: Statistical Machine Learning, CSCI 69500: MS Capstone Project or STAT 59800: Topics in Statistical Methods

c) Sample Curriculum

In the sample curriculum below, a number of electives are listed for each semester, but these are meant to be options from which the students can select. The students will select two core courses and one elective in two of the three semesters and two core courses and two electives in one of the three semesters in order to complete 30 credits over 3 semesters.

Semester 1	Core	• CSCI 59000: Introduction to Data Science • STAT 51200: Applied Regression Analysis
	Elective	• CSCI 58000: Algorithm Design, Analysis, and Implementation • STAT 52100: Statistical Computing
Semester 2	Core	• CSCI 57300: Data Mining • STAT 52900: Applied Decision Theory and Bayesian Analysis
	Elective	• CSCI 54100: Database Systems • CSCI 55200: Advanced Graphics and Visualization • CSCI 52000: Computational Methods in Analysis

		• STAT 51400: Design of Experiments • STAT 52000: Time Series and Applications • STAT 52300: Categorical Data Analysis • STAT 53600: Introduction to Survival Analysis
Semester 3	Core	• CSCI 57800: Statistical Machine Learning • CSCI 69500 MS Capstone Project ***or*** STAT 59800: Topics in Statistical Methods
	Elective	• CSCI 59000: Large-scale Machine Learning • CSCI 59000: High Performance Computing • STAT 54200: Applied Multivariate Analysis • STAT 52501: Generalized Linear Models

d) Existing courses in the proposed curriculum

Core courses

Course Title: Applied Regression Analysis
Course Number: STAT 51200
Credit Hours: 3
Description: Theoretical foundations and the practical use of linear regression models, for applications in real data analysis projects. Students are also trained to use diagnostic tools to assess the validity of regression models, and apply variable selection and model building strategies in linear regression models. Extensive use of the computer is needed to conduct data analyses and programming in R.

Course Title: Data Mining
Course Number: CSCI 57300
Credit Hours: 3
Description: Data Mining has emerged at the confluence of artificial intelligence, statistics, and databases as a technique for automatically discovering summary knowledge in large datasets. This course introduces students to the process and main techniques in data mining, including classification, clustering, and pattern mining approaches. Data mining systems and applications will also be covered, along with selected topics in current research.

Course Title: Applied Decision Theory and Bayesian Analysis
Course Number: STAT 52900
Credit Hours: 3
Description: Selected modeling/data analysis using Bayesian methodology, which includes a term project with presentation and report. Bayes Theorem, Bayesian influence, posterior distributions, multi-parameter models, hierarchical models, large sample methods, model selection.

Course Title: Statistical Machine Learning
Course Number: CSCI 57800
Credit Hours: 3
Description: This course will provide an introductory to mid-level coverage of concepts and techniques in the field of machine learning with more emphasis given on statistical machine learning. Topics to be discussed include: Generative and discriminative models for classification and regression, posterior distributions and inference, conjugate distributions, model generalizability, kernel machines, dimensionality reduction, introduction to probabilistic topic models, graphical models and belief propagation, expectation-maximization, deterministic and stochastic inference. Prerequisites: Calculus, linear algebra, probability and random variables, basic knowledge of optimization techniques, Matlab/R programming.

Course Title: MS Project
Course Number: CSCI 65900
Credit Hours: 3
Description: The student integrates and applies the knowledge gained from the formal course work to formulate and execute a solution to a problem of practical importance. The faculty advisor and the sponsoring organization mentor, if applicable, provide guidance and evaluation.

Course Title: Topics in Statistical Methods
Course Number: STAT 59800
Credit Hours: 3
Description: Directed study for students who wish to undertake individual reading and study on approved topics.

Elective courses
Course Title: Computational Methods in Analysis
Course Number: CSCI 52000
Credit Hours: 3
Description: A treatment of numerical algorithms for solving classical problems in real analysis with primary emphasis on linear and nonlinear systems of equations and on optimization problems; the writing, testing, and comparison of numerical software for solving such problems; and a discussion of the characteristics of quality software for implementing these algorithms.

Course Title: Database Systems
Course Number: CSCI 54100
Credit Hours: 3
Description: Fundamentals for the logical design of database systems. The entity-relationship model, semantic model, relational model, hierarchical model, network model. Implementations of the models. Design theory for relational databases. Design of query languages and the use of semantics for query optimization. Design and verification of integrity assertions, and security. Introduction to intelligent query processing and database machines.

Course Title: Advanced Graphics and Visualization
Course Number: CSCI 55200
Credit Hours: 3
Description: An introduction to data visualization methods and tools, and related graphics techniques. Students will explore a variety of data representation and modeling techniques, their

corresponding visualization algorithms, and practical visualization applications in scientific, engineering, and biomedical fields. Prerequisite: 55000.

Course Title: Algorithm Design, Analysis and Implementation
Course Number: CSCI 58000
Credit Hours: 3
Description: Basic techniques for designing and analyzing algorithms: dynamic programming, divide-and-conquer, balancing, upper and lower bounds on time and space costs, worst case and expected cost measures. A selection of applications such as disjoint set union/find, graph algorithms, search trees, pattern matching. The polynomial complexity classes P, NP, and co-NP; intractable problems.

Course Title: Statistical Computing
Course Number: STAT 52100
Credit Hours: 3
Description: This course demonstrates how computing can be used to understand the performance of core statistical methods and introduces modern statistical methods that require computation in their applications which includes optimization methods to find MLE, simulation and bootstrapping. We mainly cover relevant programming fundamentals in three programming environments: R, MATLAB and SAS.

Course Title: Design of Experiments
Course Number: CSCI 51400
Credit Hours: 3
Description: This course covers the principles of experimental design, the analysis of variance method, the difference between fixed and random effects and between nested and crossed effects, and the concept of confounded effects. The designs covered include completely random, randomized block, Latin squares, split-plot, factorial, fractional factorial, nested treatments and variance component analysis, etc. SAS, R, or MINITAB is the primary software used in this course.

Course Title: Time Series and Applications
Course Number: STAT 52000
Credit Hours: 3
Description: This course provides an introduction to the techniques and the theory of time series analysis; both the time domain approach and the frequency domain approach are discussed. It covers topics including smoothing methods for forecasting, modeling and forecasting using univariate, autoregressive, moving average models, ARMA models, Box-Jenkins techniques, ARIMA models, etc. Applications to engineering, economics and environmental problems are examined.

Course Title: Categorical Data Analysis
Course Number: STAT 52300
Credit Hours: 3
Description: Models generating binary and categorical response data, two-way classification tables, measures of association and agreement, goodness-of-fit tests, testing independence, large sample properties. General linear models, logistic regression, and probit and extreme value models. Loglinear models in two and higher dimensions, maximum likelihood estimation, testing goodnessof-fit, partitioning chi-square, and models for ordinal data. Model building, selection, and

diagnostics. Other related topics as time permits. Computer applications using existing statistical software.

Course Title: Introduction to Survival Analysis
Course Number: STAT 53600
Credit Hours: 3
Description: This course introduces basic concepts in survival analysis with examples of typical data sets in biomedical applications of this methodology. It includes estimation of parameters, hypothesis testing, regression analysis, and multivariate models for censored and/or truncated data.

Course Title: Large-scale Machine Learning
Course Number: CSCI 59000
Credit Hours: 3
Description: This course covers the fundamentals of distributed computing for data science and machine learning at scale. Design of parallel machine learning algorithms for regression, classification and clustering. Implementation of distributed algorithms using Python, Hadoop, and Spark.

Course Title: High Performance Computing
Course Number: CSCI 59000
Credit Hours: 3
Description: This course covers important topics of designing parallel algorithms, writing faster programs, developing large distributed applications, and performance analysis and optimization. The course will include lectures, homework, paper discussion, and a course project. As a prerequisite, students should ideally have some programming experience in C or a similar language.

Course Title: Applied Multivariate Analysis
Course Number: STAT 52400
Credit Hours: 3
Description: Extension of univariate tests in normal populations to the multivariate case, equality of covariance matrices, multivariate analysis of variance, discriminant analysis and misclassification errors, canonical correlation, principal components, and factor analysis. Strong emphasis on the use of existing computer programs.

Course Title: Generalized Linear Models
Course Number: STAT 52501
Credit Hours: 3
Description: This course focuses on generalized linear models, and it also covers AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion), generalized estimating equations, and correlated binary data analysis. Some of the specific topics it covers are model fitting, estimation, inference, normal linear models, logistic regression, Poisson regression and log-linear models, mixture models, and Markov chains.

e) Courses to be added

Core courses
Course Title: Introduction to Data Science

Course Number: CSCI 59000
Credit Hours: 3
Description: Review of linear algebra and multivariate calculus. Data manipulation, analysis, and visualization. It includes principles of statistical modeling, dimensionality reduction, and classification (supervised/unsupervised), as well as review of some elementary data structures (arrays, stacks, and queues). Emphasis is given to software technologies used in data science applications, in science, government, and industry (including Python, Java, and C++).

Elective courses
None

2) List of faculty members and administrators

CVs of faculty members who will be teaching the courses listed in the curriculum above are included as an appendix.

IUPUI

SCHOOL OF INFORMATICS
AND COMPUTING

INDIANA UNIVERSITY–PURDUE UNIVERSITY
Indianapolis

January 30, 2018

Indiana Commission for Higher
Education 101 West Ohio Street, Suite
300 Indianapolis, IN 46204-4206

Dear Commissioners,

On behalf of the Indiana University School of Informatics and Computing at IUPUI, I am writing in support of the proposed Master of Science program in Computational Data Science to be offered by the Purdue School of Science's Department of Computer and Information Science and Department of Mathematical Sciences. The program leverages the research strengths of these two departments and their existing graduate course offerings. The program's core delves deeply into the development of statistical, mathematical, and computational algorithms for data analysis and theoretical and applied platforms for data science. Their emphasis on data analysis nicely complement our existing MS in Applied Data Science in the School of Informatics and Computing, which instead focuses on applications within the breadth of the data science pipeline, from data acquisition to visualization.

Given the shortfall of data science professionals, both locally and nationally, the strategic initiatives of the campus and university, and opportunities for large external collaborative grants in data science, this program will enable IUPUI, as the first campus in Indiana with a Data Science Ph.D., to reach critical mass in both teaching and research in this highly valued field. The new program will enhance our collaborations with the School of Science. We strongly support the request for a MS in Computational Data Science, which addresses the significant need for trained data science professionals in central Indiana and beyond.

Sincerely,

Mathew J. Palakal, Ph.D.
Executive Associate Dean IU
School of Informatics and
Computing IUPUI

IUPUI

DEPARTMENT OF COMPUTER INFORMATION AND GRAPHICS TECHNOLOGY

SCHOOL OF ENGINEERING AND TECHNOLOGY
Indiana University–Purdue University
Indianapolis

Indiana University Graduate School
University Library, Room 1170 755
West Michigan Street
Indianapolis, IN 46202

December 12, 2017

Re: New Degree preproposal for Master of Science in Data Science from Departments of Computer & Information Science and Mathematical Sciences;

Dear Graduate Affairs Committee,

The Departments of Computer & Information Science and Mathematical Sciences has proposed the new degree of Master of Science in Data Science. After reviewing the preproposal provided by CSCI and Math, I am writing to provide my strong support in this new degree proposal. Our support is mainly based on two observations on the new degree:

First, this program will utilize existing graduate courses offered by the two Departments, with related, complimentary topics in Computer Science and Statistics. The CSCI and Math departments has strong faculty expertise in the area of data science. The existing course included in the proposal demonstrated both depth and breadth of graduate education in the data science areas. The new degree program also has in a carefully designed sequence that incrementally yields deeper understanding of different theoretical and applied platforms of data science.

Second, the program will prepare students to enter the expanding workforce in the data science field by providing a solid grounding in the corresponding computational and mathematical aspects, from the fundamentals to advanced topics. This will make the IUPUI students competitive in the job market in Data Science. And that will help to increase the impact of IUPUI's graduate program in both local and national economy. The graduate students from this rigorous program are expected to be competitive applications to the leading employers in Indiana, such as the Eli Lily company.

Based on these two points, the CIGT program supports this proposal and the plan for the new degree. The CIGT program also expects to have collaborations with the newly proposed data science program from the applied data analytics side.

Sincerely,

Feng Li
Interim Chair and Associate Professor
Department of Computer Information and Graphics Technology
Purdue School of Engineering and Technology
IUPUI, 317-278-8391

IUPUI

RICHARD M. FAIRBANKS
SCHOOL OF PUBLIC HEALTH

INDIANA UNIVERSITY
Indianapolis

December 19, 2017

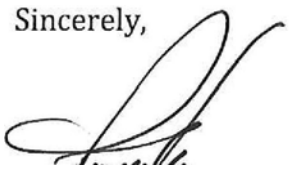
Shiaofen Fang, PhD,
Professor and Chair, Computer Science
Department of Computer and Information Science
IUPUI

Dear Dr. Fang,

As Founding Dean of the IU Richard M. Fairbanks School of Public Health, I am providing a letter of support for your proposed MS degree, however, I would like to recommend more specificity in the degree title. Would you be willing to consider (1) Computational Data Science, (2) Theoretical Data Science or (3) Mathematical Data Science? The title of Data Science is very broad and conjures up different things to different people. Just as our school titled our new undergraduate degree Health Data Science, we recommend that you be more specific to avoid confusion among students.

In public health, we are acutely aware of the role that strong data science research and practice plays in producing quality data analysis, the results of which provide the very basis upon which public and population health decisions are made. Biostatistics with data science is becoming increasingly important because current problems increasingly require very large data sets (e.g,, electronic health-record data). Biostatistical methods are designed to accommodate these large quantities of data, and practitioners need data science skills to manage them. Given Indiana's emerging interest and need for educational programs in this area, I support your proposal for an MS with the caveat that a narrower title be given to the degree.

Sincerely,



Paul K alverson, DrPH, FACHE
Foun • ng Dean and Professor

IUPUI

DEPARTMENT OF
ELECTRICAL AND
COMPUTER ENGINEERING

SCHOOL OF ENGINEERING AND TECHNOLOGY
Indiana University–Purdue University
Indianapolis

January 30, 2018

To whom it may concern,

This letter is a letter of support for the MS in Computational Data Science that has been proposed by School of Science at Indiana University Purdue University Indianapolis. This program is a well thought-out curriculum for the computational data science field. The proposed curriculum combines the graduate statistic courses that are the most relevant for this field with data oriented graduate computer science courses forming a strong, focused MS degree.

The Electrical & Computer Engineering department at Indiana University Purdue University Indianapolis supports this curriculum initiative, as it represents a solid curriculum, which will address central Indiana industry needs within the computational data science field.

Sincerely,

Brian King
Associate Professor of Electrical & Computer Engineering
Chair, Electrical & Computer Engineering
Department of Electrical and Computer Engineering
Indiana University - Purdue University Indianapolis
723 West Michigan Street, SL160
Indianapolis, IN 46202
Phone: +1 317 274 9723

PURDUE
UNIVERSITY

DEPARTMENT OF COMPUTER SCIENCE

Sunil Prabhakar
Director, Integrative Data Science Initiative

July 25, 2018

Prof. Shiaofen Fang
Head, Department of Computer and Information Science
Indiana University-Purdue University, Indianapolis

Dear Prof. Fang,

As Director of Purdue's Integrative Data Science Initiative, I am pleased to provide this letter of support for your proposed Master of Science in Computational Data Science degree program offered jointly by the Computer and Information Science department in collaboration and the department of Statistics at IUPUI.

The proposal is excellent, timely, and responds to the growing need for well-trained data scientists in Indiana and beyond. I am especially pleased to see that the program aims to provide a solid, comprehensive background in both the foundational theory and applications. The intended coverage of theoretical, statistical, and computational aspects of the program, and the goal of teaching a deeper understanding of the methods and techniques rather than a "black box" approach of using existing tools to apply machine learning methods is judicious.

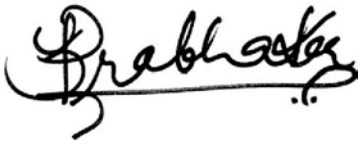
The emphasis on health-oriented programs is very likely to be a great fit for IUPUI students given the existing strengths at IUPUI in healthcare as evidenced by the School of Medicine and School of Public Health. The program is well designed and the faculty in both departments are well positioned to create a program that will be in great demand from students.

In recognition of the importance of Data Science, Purdue recently launched the Integrative Data Science Initiative (IDSI) to support and coordinate our efforts in education and research in this emerging area. Your proposed degree program aligns well with our vision of data science, and will play a complementary role to our efforts in helping address the looming lack of well-trained data scientists in the workforce. I look forward to identify opportunities for collaboration between IDSI and your program.

I enthusiastically support your proposed degree program.

305 N. UNIVERSITY STREET • WEST LAFAYETTE, IN 47907-2107
(765) 494-6010 • FAX: (765) 494-0739

Sincerely,



Sunil Prabhakar
Purdue University
Director, Integrative Data Science Initiative
Professor of Computer Science

Tyler Foxworthy

10 West Market Street, Suite
1950 Indianapolis, Indiana
46202 317-701-3703
tyler@demandjump.com

May 16, 2018

To Whom It May Concern,

I would like to extend my sincere support for the creation of a new Master of Science program in Computational Data Science at IUPUI. As a practicing data scientist and entrepreneur, I have been generally disappointed with the availability of qualified early career data scientists in the Midwest. Many of those candidates whom I interview for data science and machine learning roles lack a rigorous understanding of both the theoretical and practical aspects of applying computational and statistical techniques to solving real world problems. There has unfortunately been a proliferation in the tech industry of "black box" data scientists, who apply machine learning techniques to problems without even a cursory understanding of their mechanics and limitations. This has many dangerous implications for both the viability of this field and poses a danger to investors in technology companies at large. Rigorous education and preparation at the university level is the first and most critical hedge against this risk.

I've had the pleasure of getting to know many of the faculty in both the Computer Science and Mathematics departments at IUPUI when I've visited to deliver guest lectures and seminars over the last few years. I've been wholly impressed with the quality of faculty and students in these programs, and am confident that the proposed degree program will be both immensely popular and effective in its implementation. I very much hope to be able to lend my ongoing support to this proposed degree program and see it as a great benefit to not only my business, but to the greater tech community within the State of Indiana.

Sincerely,

Tyler Foxworthy

Chief Scientist, DemandJump Inc.

SCHOOL OF INFORMATICS,
COMPUTING, AND ENGINEERING
INDIANA UNIVERSITY

May 12, 2018

Dear colleagues,

As Chair of the department of Computer Science in the School of Informatics, Computing, and Engineering at Indiana University, Bloomington, I write in strong support of the Master of Science proposal in Computational Data Science submitted by the departments of Computer & Information Science and Mathematical Sciences of the IUPUI School of Science.

In short, there is a clear demand for Data Science skills in the state of Indiana. While our Bloomington school is o↵ering both online and residential programs in Data Science, a complementary residential o↵ering in the Indianapolis area is well-suited to the residential demand in that area.

Sincerely,

Amr A. Sabry
Professor and Chair of
Computer Science Department

700 N. WOODLAWN AVE, LUDDY HALL, BLOOMINGTON, INDIANA
47408-3901 • PHONE (812) 855-6486 • FAx (812) 855-4829

Members of the Graduate Council,

Please accept this letter as support for the new Master of Science Program in Computational Data Science.

I would expect other companies are in a similar situation as Eli Lilly, where we find the demand for highly qualified data science professionals to far exceed the available pool of talent in Central Indiana. Being highly qualified in this area implies cross-disciplined knowledge of one or more business areas, data management, analytics capabilities, programming, and statistics. A program that offers both the theory and the application in these domains that can be translated into real-world scenarios is definitely needed.

These skills are critical for the emerging cognitive era, where data driven insights will drive businesses and create new opportunities for competitive advantage. At Eli Lilly, we find elements of advanced analytics and data sciences appearing in both unique internally developed use cases across all lines of business (e.g. research, clinical, manufacturing, marketing, digital health, etc.), and also as evolving components in a wide variety of industry platforms we utilize. This is not a short-term need, but reflective of a new way of doing business, where data is leveraged as a strategic asset, analytics drives new insights, and humans work alongside machines in a complimentary manner. In a world where there will never be less data, but only exponential growth, these types of programs offer the skills needed to discover, understand, and apply new insights.

Evidence of the organizational impact can be seen in our recent decisions, including the hiring of a new Chief Data and Analytics Officer and the establishment of an Advanced Analytics and Data Sciences organization. Both are reflective of our commitment to grow and lead the business into the cognitive era.

By offering this new degree, you create a unique opportunity to meet the much needed skills of companies throughout the State.

Sincerely,

Kent Supancik
Senior Director, Enterprise IT Architecture and Information Services
Eli Lilly and Company

BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: Benzion Boukai

eRA COMMONS USER NAME (credential, e.g., agency login): bboukai

POSITION TITLE: Chancellor's Professor of Statistics

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.)*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	Completion Date MM/YYYY	FIELD OF STUDY
University of Haifa, Haifa, Israel	B.A.	10/1983	Statistics
University of Haifa, Haifa, Israel	M.S.	01/1985	Mathematics/Statistics
SUNY Binghamton, NY, USA	Ph.D.	10/1988	Statistics

Academic Appointments:

1988-1990	Visiting Assistant Professor, Department of Statistics, Purdue University
1990-1996	Assistant Professor, Department of Mathematical Sciences, IUPUI
1996-2000	Tenured Associate Professor, Department of Mathematical Sciences, IUPUI
1997-1998	Visiting Associate Professor, University of Haifa, Israel
1999-2009	Chair, Department of Mathematical Sciences, IUPUI
Spring 2012	Visiting Professor, University of Haifa, Israel
2000- present	Professor, Department of Mathematical Sciences, IUPUI
2017-present	Chancellor's Professor, Department of Mathematical Sciences, IUPUI

P

1997-2003	Associate Editor for the Journal of Statistical Planning and Inference
2002-2006	Associate Editor for Communication in Statistics–**B**–Computations and Simulations
2002-2006	Associate Editor for Communication in Statistics--**A**--Theory and Methods
2007-2011	Director, Signature Center for Mathematical Biosciences, CMB
11/13 & 3/15	Visiting Professor, New York University-Shanghai, China
2003-present	Associate Editor for Sequential Analysis
2007- present	Co-Director, Biostatistics PhD program, IUPUI
2017- present	Member of the IUPUI Forensic Science Research Cluster

Honors and Awards:

1988	Distinguished Dissertation Award, SUNY Binghamton
1997	Teaching Excellence Recognition Award (TERA) at IUPUI
1999	Teaching Excellence Recognition Award (TERA) at IUPUI
4/2017	Chancellor's Professorship Award, IUPUI

MURAT DUNDAR
SL 280C, 723 W. Michigan St., Indianapolis, IN 46202
Tel: (+1) 317 278 6488 Fax: (+1) 317 274 9742 Email: mdundar@iupui.edu

Dr. Dundar's area of expertise is in machine learning and data mining with a focus on non-parametric Bayesian models and inference, learning with partially-observed data, online and offline class discovery and modeling, semi-supervised, multi-instance, and multi-task learning. His research is mainly driven by real-world problems in hyperspectral image analysis and remote sensing, computer aided diagnosis/detection, bio-detection, and flow cytometry data analysis.

EDUCATION:
Ph. D. (2003) and M.S. (1999) in Electrical and Computer Engineering, Purdue University
B.S. (1997) in Electrical and Electronics Engineering, Bogazici University

PROFESSIONAL EXPERIENCE:
Associate Professor, Indiana University – Purdue University, Indianapolis, 2014-present
Assistant Professor, Indiana University – Purdue University, Indianapolis, 2008-2014
Research Scientist, Siemens Medical Solutions, 2003-2008

AWARDS, PROFESSIONAL SERVICE, AND SOCIETIES
NSF CAREER Award, 2013. Best paper award (International Conference on Pattern Recognition), 2010. Data mining practice prize (ACM SIGKDD), 2009. Reviewer for IEEE TGARS, TPAMI, TIP, TMI, SMC. PC member for ICML, NIPS, AAAI, AIStats, IEEE ICDM, ACM SIGKDD, and SIAM SDM. Reviewer for NIH BMIT-B, NSF IIS, NSF SBIR/STTR panels. Member, IEEE and ACM SIGKDD.

SELECT RECENT PUBLICATIONS:

Halid Z. Yerebakan and Murat Dundar. "Partially collapsed parallel Gibbs sampler for Dirichlet process mixture models." *Pattern Recognition Letters* 90, 22-27, 2017.

Rajwa, Bartek, et al. "Automated assessment of disease progression in acute myeloid leukemia by probabilistic analysis of flow cytometry data." *IEEE Transactions on Biomedical Engineering* 64.5, 1089-1098, 2017.

Halid Z. Yerebakan, Bartek Rajwa, Murat Dundar, "The Infinite Mixture of Infinite Gaussian Mixtures," in *Proceedings of the Advances on Neural Information Processing Systems (NIPS'14)*, Montreal, CA, Dec 2014.

Murat Dundar, Ferit Akova, Halid Z. Yerebakan, Bartek Rajwa, "A Non-parametric Bayesian Model for Joint Cell Clustering and Cluster Matching: Identification of Anomalous Sample Phenotypes with Random Effects," *BMC Bioinformatics* 15 (1), 314, 2014.

RECENT GRANTS:

NSF, CAREER: Self-adjusting Models as a New Direction in Machine Learning, 3/1/13-2/28/18, $499K. NIH, Automated Spectral Data Transformations and Analysis Pipeline for High-throughput Flow Cytometry, 7/1/12-6/30/14, Co-I, $110K.

NIH, Machine-Learning Approach to Label-free Detection of new Bacterial Pathogens, 5/1/10-4/29/12, PI, $190K.

Shiaofen Fang

723 W. Michigan St., SL 280 317-274-9731
Indianapolis, IN 46202 E-mail: sfang@cs.iupui.edu

EDUCATION

- Ph.D, Computer Science, University of Utah, 1992.
- M.S., Applied Mathematics, Zhejiang University, China, 1986.
- B.S. Mathematics, Zhejiang University, China, 1983.

A P P O I N T M E N T S :

- 2007 – Present: Chair, Department of Computer and Information Science (CIS), Indiana University Purdue University Indianapolis (IUPUI)
- 2009 – Present : Professor, CIS Department, IUPUI
- 2002 –2009: Associate Professor, CIS Department, IUPUI
- 1996 –2002: Assistant Professor, CIS Department, IUPUI
- 1993 –1996: Research Staff, ISS, National University of Singapore.
- 1992 –1993: Assistant Professor, CAD program, School of Architecture, The Ohio State University.

RECENT EXTERNAL RESEARCH GRANTS

1. Integrative Bioinformatics Approaches to Human Brain Genomics and Connectomics. NIH-NIBIB, R01 EB022574. $1,943,717. Co-PI (PI: Li Shen). 8/1/16 – 4/30/20
2. Health-Terrain: Visualizing Large Scale Health Data, PI, DoD (Army), $661,035, 3/1/13 – 9/30/15.
3. 3D Facial Imaging on FASD , co-PI (PI: Tatiana Foroud), NIH, $1,500,000, 06/01/08 – 05/31/13.
4. Mouse Model Neuro-Facial Dysmorphology: Translational and Treatment Studies, co-PI (PI: Feng Zhou), NIH, $1,200,000, 06/01/08 – 05/31/13.
5. A Cross-Cultural Longitudinal Assessment of FASD, Co-PI (PI: Tatiana Foroud), NIH, $662,733, 9/29/03 – 9/28/07.

PUBLICATIONS

1. Li H, Fang S, Contreras JA, West JD, Risacher SL, Wang Y, Sporns O, Saykin AJ, Goni J, Shen L. Brain explorer for connectomic analysis. Brain Informatics, 2017.
2. Mao, Songan; Wu, Huanmei; Sandison, George; Fang, Shiaofen, Iterative volume morphing and learning for mobile tumor based on 4DCT. Phys. Med. Biol. 62 (2017) 1501–1517.
3. Liangchen Hu, Huahao Shou and Shiaofen Fang. A PIA optimization algorithm for non-uniform B-spline curves and surfaces. International Journal of Modelling and Simulation 37(3), 2017, 167-177
4. Savitha Baskaran, Shiaofen Fang, and Shenghui Jiang. Spatiotemporal Visualization of Traffic Paths Using Color Space Time Curve. Big Spatial Data Workshop, IEEE Big Data 2017 (Best Paper Award).
5. Keerthika Koka, Shiaofen Fang. Online Review Analysis by Visual Feature Selection. 3rd IEEE International Conference on Big Data Intelligence and Computing (IEEE DataCom) 2017
6. Huang Li, Shiaofen Fang, Bob Zigon, et al. Olaf Sporns, Andrew Saykin, Joaquin Goni, and Li Shen. BECA: A Software Tool for Integrated Visualization of Human Brain Data. Workshop on Big Data and Visualization for Brainsmatics, Brain Informatics, Beijing, 2017
7. Wang, Jiachen; Fang, Shiaofen; Li, Huang; Go–i, Joaqu'n; Saykin, Andrew J.; Shen, Li. Multigraph Visualization for Feature Classification of Brain Network Data. EuroVis Workshop on Visual Analytics (EuroVA), pp.61-65, 2016
8. Shenghui Jiang, Shiaofen Fang, Shaun Grannis. Spiral Theme Plot, EuroVis'16, 109-122

Mohammad Al Hasan

Department of Computer Science, IUPUI
723 W. Michigan St., SL 277
Indianapolis, IN 46226

Phone: 317 274-3862
alhasan@iupui.edu
http://www.cs.iupui.edu/~alhasan

Professional Preparation

Ph.D., Computer Science	2009	Rensselaer Polytechnic Institute, Troy, NY
MS, Computer Science	2002	University of Minnesota, Minneapolis, MN
BSc (Engg.), Computer Sc. & Engg.	1998	Bangladesh Univ. of Engg. & Tech. (BUET), Dhaka

Professional Experience

07/2016–Present	Associate Professor (with tenure), Computer Science, IUPUI, Indianapolis, IN
08/2010–06/2015	Assistant Professor (tenure-track), Computer Science, IUPUI, Indianapolis, IN
07/2009–08/2010	Senior Research Scientist, eBay Research Labs, San Jose, CA
05/2007–08/2007	Research Intern, IBM Almaden Research Center, San Jose, CA
05/2006–08/2006	Research Intern, IBM Almaden Research Center, San Jose, CA
08/2004–06/2009	Research/Teaching Assistant, Rensselaer Polytechnic Institute, Troy, NY

Synergistic Activities

1. NSF Panelist in IIS, BigData, SBIR and STTR division (2010, 2012, 2014, 2015, 2016, 2017, 2018)
2. Program Committee Co-chair, BigData Workshop on graph data management, and mining, (2014-2018); 10th Intl. Workshop in Data Mining in Bioinformatics (BIOKDD), 2011
3. Treasurer and Local Chair, CIKM 2016; Publicity Chair, SIAM Data Mining (SDM) 2014; Poster Chair, BigData 2016; Big Data Cup Chair 2018
4. Conference program Committees (2018): KDD, ICDM, CIKM, ECML-PKDD, BigData
5. Conference program Committees (2017): KDD, CIKM, ECML-PKDD, ICDM, BigData
6. Journal Guest Editor: Springer Data Science and Engineering Journal
7. Journal Reviewer: Knowledge and Information Systems (KAIS); VLDB Journal; IEEE Transactions on Knowledge and Data Engg. (TKDE);ACM Transactions on Knowledge and Data Discovery (TKDD); Data Mining and Knowledge Discovery (DMKD); Journal of Machine Learning Research; IEEE Transactions on Neural Networks

Publications

See in http://dmgroup.cs.iupui.edu/publication.php)

Patents

- US patent No:7881937, Method for Analyzing Patent Claims. Assignee, IBM Corporation
- US patent No:8954422,Query Suggestions for eCommerce Sites, Assignee, eBay Inc

Honors

- CareerBuilder Research Award, 2016
- NSF CAREER Award, 2012
- ACM SIGKDD Doctoral Dissertation Award, 2010
- Best Paper award, PAKDD Conference, 2009

CURRICULUM VITAE of

Fang Li

CONTACT INFORMATION

Science Building, LD 270F
Department of Mathematical Sciences, IUPUI
402 N. Blackford Street
Indianapolis, IN 46202-3216

Tel: (317) 274-6938 (O)
Fax: (317) 274-3460
Email: fli@math.iupui.edu

EDUCATION

- Ph.D., Statistical Science (2004), Michigan State University, East Lansing, MI
- M.S., Statistical Science (1998), Beijing Normal University, Beijing, P. R. China
- B.S., Statistics & Probability (1995),Beijing Normal University, Beijing, P. R. China

PROFESSIONAL EXPERIENCE

- Associate Professor, Department of Mathematical Sciences, Indiana University-Purdue University Indianapolis, 07/2010-present
- Visiting Professor, Sun-Yat-Sen University, China, 07/2111-12//2011
- Assistant Professor, Department of Mathematical Sciences, Indiana University-Purdue University Indianapolis, 08/2004-06/2010

FUNDING

1. Li, F., Co-PI, (2014), "Neighborhood poverty and sexually transmitted infections." NIAID/NIH, 02/01/2014-12/30/2014, $9337.84, funded
2. Li,. F., Key Contributor, (2014), "Teladoc consults Forecasting." Teladoc, Inc., 01/01/2014-12/31/2014, $8145.15, funded (PI: Dr. Julie Meek)
3. Li, F., Project Director (2010), "Spatial Modelling of the Indiana State Traffic Data," Indiana Department of Transportation, 06/01/2010 – 07/31/2011, $65,550, funded

PUBLICATIONS

1. **Li, F.** (2014) Chapter 15: Comparison of Autoregressive Curves through Partial Sums of Quasi-residuals. *Contemporary Developments in Statistical Theory*, Springer International publishing Switzerland.
2. Tung, C.Y., Lewis, D., Han, L., Jaja, M., Yao, S., **Li, F.**, Robertson, M.J., Zhou, B., Sun, J. and Chang, H.C. (2014) Activation of dendritic cell function by soypeptide lunasin as a novel vaccine adjuvant, *Vaccine*, DOI: 10.1016/j.vaccine.2014.07.103
3. King, J.L., Knapp, K.E., Shaikh, A., **Li, F.**, Sabau, D., Pascuzzi, R.M., Osburn, L.L. (2017) Cortical Activity Changes after Art Making and Rote Motor Movement as Measured by EEG: A Preliminary Study. *Biomedical Journal of Scientific & Technical Research (BJSTR)* Vol 1, Sept 2017 **DOI:** 10.26717/BJSTR.2017.01.000366

George Owen Mohler

gmohler@iupui.edu

PROFESSIONAL PREPARATION

University of California Santa Barbara
Ph.D. Mathematics, 2008

Indiana University
B.S. Mathematics, 2003

ACADEMIC POSITIONS

Associate Professor
Department of Computer and Information Science, Indiana University Purdue University Indianapolis
2016-present

Assistant Professor
Department of Mathematics and Computer Science, Santa Clara University 2010-2016

CAM Assistant Adjunct Professor
Department of Mathematics, UCLA
2008-2010

SYNERGISTIC ACTIVITIES

- 10 years of experience supervising undergraduate research in computational mathematics, statistics and data science. PI for NSF REU Site: Data science of risk and human activity.
- Developed predictive policing software for the Los Angeles and Santa Cruz police departments for the allocation of daily patrol resources. The project was featured in the New York Times, TIME, the Economist, Popular Science, NBC, CBS and ABC national news, and other international news outlets. This project is featured through the NSF sponsored Science Behind The News series.
- Ranked as high as 203/47438 as a Kaggle competitor and participated in 12 machine learning competitions, 3 on teams with undergraduate students. Tied for 1st place in the large business category of the 2017 NIJ Real-time Crime Forecasting Challenge.
- Reviewer for the following journals: Journal of the Royal Statistical Society A & B, Multiscale Modeling and Simulation, Annals of Applied Statistics, Annals of Statistics, SIAM J. on Applied Mathematics, International Journal of Forecasting, Nonlinear Processes in Geophysics, Applied Mathematics Research Express, Canadian Journal of Criminology and Criminal Justice, Journal of Quantitative Criminology, Discrete and Continuous Dynamical Systems. On the program committee for SocialSens 2018 and co-organized the 2016 KDD workshop on Machine Learning for Large Scale Transportation Systems.
- Signficant industrial experience. Served as Director of Data Science at Metro-mile and co-founded PredPol, a law enforcement focused predictive analytics company.

Dr. Snehasis Mukhopadhyay
Department of Computer Science
Indiana University Purdue University Indianapolis
Email: smukhopa@cs.iupui.edu

Professional Preparation

- B.E. in Electronics and Telecommunications, Jadavpur University, India, 1981 - 1985
- M.E. in Systems Science and Automation, Indian Institute of Science, Bangalore, India, 1985 - 1987
- M.S. in Electrical Engineering, Yale University, New Haven, CT, 1987 - 1991
- Ph.D. in Electrical Engineering, Yale University, New Haven, CT, 1987 – 1994

Selected Appointments

- July 2010-Present: Professor, Computer & Information Science, IUPUI; co-director, Institute of Mathematical Modeling and Computational Science, IUPUI.
- 2000 - 2006, Associate Director (Bioinformatics), School of Informatics, Indiana University
- 1993 - 1994, Staff Fellow, Analytic Processes, GM NAO R&D Center, Warren, Michigan

R

1. Piemonti, A.D., Babbar-Sebens, M., Mukhopadhyay, S. and Kleinberg, A., 2017. Interactive genetic algorithm for user-centered design of distributed conservation practices in a watershed: An examination of user preferences in objective space and user behavior. *Water Resources Research*, *53*(5), pp.4303-4326.
2. Babbar-Sebens, M., Mukhopadhyay, S., Singh, V.B. and Piemonti, A.D., 2015. A web-based software tool for participatory optimization of conservation practices in watersheds. *Environmental Modelling & Software*, *69*, pp.111-127.
3. T. Dugan, S. Mukhopadhyay, A. E. Carroll, and S. M. Downs. Machine Learning Techniques for the Prediction of Early Childhood Obesity. pp. 506-520, Applied Clinical Informatics Journal, vol. 6, issue 3, 2015..
4. Tilak, O., Martin, R. and Mukhopadhyay, S., 2011. Decentralized indirect methods for learning automata games. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *41*(5), pp.1213-1223.
5. Tilak, O. and Mukhopadhyay, S., 2011. Partially decentralized reinforcement learning in finite, multi-agent Markov decision processes. *AI Communications*, *24*(4), pp.293-309.

Synergistic Activities

1. General Chair, ACM International Conference on Information and Knowledge Management (CIKM), 2016
2. Invited Panelist, “Artificial Intelligence Panel: Is Indianapolis Ready to Become the A.I. Leader In the Midwest?”, Indiana IT Symposium, Indianapolis, IN, 2017.
3. Member, IEEE Technical Committee on Soft Computing, 2012-present
4. Proposal Reviewer and Invited Review Panelist for the National Science Foundation, 2004, 2009, 2011, 2018
5. NSF CAREER Award winner in 1996. Research continuously externally supported over past 20 years by funding agencies including NSF, NIH, and NOAA.

Xia Ning

723 W. Michigan St., SL 280 317-278-3784
Indianapolis, IN 46202 E-mail: xning@iupui.edu

EDUCATION

- Ph.D, Computer Science and Engineering, University of Minnesota, Twin Cities, US, 2012
- B.S., Computer Science, Zhejiang University, China, 2005.

APPOINTMENTS:

- 2014 – current: Assistant Professor, CIS Department, IUPUI
- 2015 – current: Member, Center for Computational Biology and Bioinformatics, IU School of Medicine
- 2017 – current: Affiliated Research Scientist, Center for Biomedical Informatics, Regenstrief
- 2012 – 2014: Research Staff Member, NEC Labs America, NJ

RECENT EXTERNAL RESEARCH GRANTS

1. CRII:III:Computational Methods to Explore Big Bioassay Data for Better Compound Prioritization. NSF IIS-1566219. $172,350. PI. 05/2016 – 04/2019.
2. SCH: INT: Mining High Dimensional Drug Interaction Induced Adverse Effects from Health Record Databases. NSF IIS-1622526. $1,149,796. IUPUI campus PI. 08/2016 – 07/2020.

PUBLICATIONS

1. Junfeng Liu and Xia Ning. Differential compound prioritization via bi-directional selectivity push with power. Journal of Chemical Information and Modeling, 57(12):2958–2975, 2017. PMID: 29178784.
2. Junfeng Liu and Xia Ning. Multi-assay-based compound prioritization via assistance utilization: A machine learning framework. Journal of Chemical Information and Modeling, 57(3):484–498, 2017. PMID: 28234477.
3. Chien-Wei Chiang, Pengyue Zhang, Xueying Wang, Lei Wang, Shijun Zhang, Xia Ning, Li Shen, Sara K. Quinney, and Lang Li. Translational high-dimensional drug interaction discovery and validation using health record databases and pharmacokinetics models. Clinical Pharmacology & Therapeutics, 2017
4. Xia Ning, Titus Schleyer, Li Shen, and Lang Li. Pattern discovery from directional high-order drug-drug interaction relations. In 2017 IEEE International Conference on Healthcare Informatics, ICHI'17.
5. Xiao Bian, Feng Li, and Xia Ning. Kernelized sparse self-representation for clustering and recommendation. In Proceedings of the 2016 SIAM International Conference on Data Mining, SDM'16, pages 10–17, 2016

Fengguang Song
Assistant Professor
Department of Computer and Information Science
Indiana University-Purdue University Indianapolis (IUPUI)
Web: http://www.cs.iupui.edu/~fgsong
E-mail: fgsong@cs.iupui.edu

a. Professional Preparation

- Postdoc Research Associate, Innovative Computing Laboratory, University of Tennessee, 2010-2012.
- Ph.D. in Computer Science, University of Tennessee at Knoxville, TN, 2009.
- M.S. in Computer Science, University of British Columbia, Vancouver, Canada, 2002.
- B.S. in Computer Science, Zhengzhou University, Zhengzhou, China, 1996.

b. Appointments

- CIS Department, IUPUI, Indianapolis, IN: Assistant Professor, 2013-Present.
- Computer Science Lab, Samsung Research America-Silicon Valley, California: Senior Researcher, 2012-2013.

c. General Summary

Dr. Song's expertise is in advanced parallel architecture, linear algebra, and high performance computing. He currently conducts research at the frontiers of computer science towards exascale computing and big data science discovery across different disciplines. In particular, he focuses on parallel algorithms, software, and advanced architectures for scientific computing, life science, simulation, and knowledge discovery. He designs innovative algorithms and software systems that can scale on large high-end systems with heterogeneous many-cores and accelerators at extreme scales.

d. Selected Publications

[SC'17] X. Liang, J. Chen, D. Tao, S. Li, P. Wu, H. Li, K. Ouyang, Y. Liu, **F. Song**, Z. Chen, "Correcting Soft Errors Online in Fast Fourier Transform", ACM/IEEE Conference on Supercomputing (SC'17), Denver, CO, November 2017.

[ICS'14] Fengguang Song and Jack Dongarra, "Scaling Up Matrix Computations on Shared-Memory Manycore Systems with 1000 CPU Cores", The 28th ACM International Conference on Supercomputing, Munich, Germany, June 2014.

[ICS'12] Fengguang Song, Stanimire Tomov, Jack Dongarra, "Enabling and Scaling Matrix Computations on Heterogeneous Multi-Core and Multi-GPU Systems", The 26th ACM International Conference on Supercomputing, San Servolo Island, Venice, Italy, June 2012.

[SC'10] Fengguang Song, Hatem Ltaief, Bilie Hadri, Jack Dongarra, "Scalable Tile Communication-Avoiding QR Factorization on Multicore Cluster Systems", ACM/IEEE Conference on Supercomputing (SC'10), New Orleans, LA, November 2010.

[SC'09] Fengguang Song, Asim YarKhan, Jack Dongarra, "Dynamic Task Scheduling for Linear Algebra Algorithms on Distributed-Memory Multicore Systems", ACM/IEEE Conference on Super- computing, Portland, OR, November 2009.

[HPDC'07] Song, F., Moore, S., Dongarra, J., "Feedback-Directed Thread Scheduling with Memory Considerations", Sixteenth IEEE International Symposium on High-Performance Distributed Computing, Monterey Bay, CA, June 2007.

e. Current Grants

i. Co-PI, A New 3D Parallel Immersed Boundary Method with Application to Hemodialysis (NSF Award# 1522554 – $209,314.00/3yrs; my share: $90,039.00; 09/15/2015 to 09/14/2018), National Science Foundation.

ii. NSF Jetstream project, 12/01/2014 – 11/30/2019, $6,576,101.00, Senior Investigator, National Science Foundation.

E N

Curriculum Vitae

EDUCATION

PhD, Statistics	Florida State University, Tallahassee, FL	08/2003 - 12/2007
MS, Statistics	Florida State University, Tallahassee, FL	08/2003 - 12/2005
BS, Mathematics	Nanjing University, Nanjing, P.R.China	09/1997 - 07/2001

EMPLOYMENT

Associate Professor	Department of Mathematical Sciences, IUPUI	08/2016 - present
Assistant Professor	Department of Mathematical Sciences, IUPUI	08/2010 - 07/2016
Assistant Professor	Institute of Public Health, College of Pharmacy, Florida A&M University	01/2008 - 07/2010

SELECTED PUBLICATIONS Peer-Reviewed Papers

1. Peng H. and Tan F. (2018) Jackknife Empirical Likelihood Goodness-Of-Fit Tests For U-Statistics Based General Estimating Equations. Bernoulli 24(1), 449 — 464
2. Dutton G., Gowey M., Tan F., Zhou D., Ard J., Perri M., Lewis C. (2017) Comparison of an Alternative Schedule of Extended Care Contacts to a Self-Directed Control: A Randomized Trial of Weight Loss Maintenance. International Journal of Behavioral Nutrition and Physical Activity 14:107 DOI 10.1186/s12966-017-0564-1
3. Tan F., Tang Y., Peng H., (2015). The Multivariate Slash and Skew-Slash Student t Distributions. Journal of Statistical Distributions and Applications. 2:3 doi:10.1186/s40488-015-0025-9. http://www.jsdajournal.com/content/2/1/3
4. Xiao H., Tan F., Adunlin G., Ali A., Goovaerts P., Gwede C.K., Huang Y. (2015) Factors Associated with Overall Survival Prostate Cancer in Florida: A Multilevel Analysis. Journal of Health Care for the Poor and Underserved. 26(1):266-77. doi: 10.1353/hpu.2015.0007. http://muse.jhu.edu/journals/journal_of_health_care_for_the_poor_and_underserved/v026/26.1.xiao.html
5. Xiao H., Tan F., Goovaerts P., Ali A., Adunlin G., Gwede C.K., Huang Y. (2014) Impact of Comorbidities on Prostate Cancer Stage at Diagnosis in Florida. American Journal of Men's Health, doi: 10.1177/1557988314564593 http://www.ncbi.nlm.nih.gov/pubmed/25542838
6. Xiao H., Tan F., Goovaerts P., Adunlin G., Ali A., Huang Y., Gwede C.K. (2013). Factors Associated with Time-to-Treatment of Prostate Cancer in Florida. Journal of Health Care for the Poor and Underserved, 24 (2013): 132-146. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4337853/
7. Tan F., Rayner G., Wang X. and Peng H. (2010). A Full Likelihood Procedure of Exchangeable Negative Binomials for Modelling Correlated and Overdispersed Count Data. Journal of Statistical Planning and Inference, Volume 140, Issue 10, 2849-2859. http://www.math.iupui.edu/~hpeng/ enb-0907.pdf
8. Dutton G., Tan F., Perri M., Stine C., Dancer-Brown M., Goble M. and Van Vessem N. (2010). What Words Should We Use When Discussing Excess Weight? Journal of the American Board of Family Medicine, 23: 606-613. http://www.jabfm.org/content/23/5/606.long
9. Dutton, G., Tan, F., Provost, B., Allen, B., Sorenson, J., Smith, D. (2009). Relationship between self-efficacy and physical activity among patients with type 2 diabetes. Journal of Behavioral Medicine, 32(3), 270-277. http://www.ncbi.nlm.nih.gov/pubmed/19156510

SELECTED TEACHING

STAT 536	Introduction to Survival Analysis	Grad stats/biostats
STAT 52501	Generalized Linear Models	Grad stats/biostats
STAT 521	Statistical Computing	Grad stats/biostats
STAT 512	Applied Regression Analysis	Grad stats/biostats
STAT 511	Statistical Methods	Grad non-stats majors
STAT 350	Introduction to Statistics	Undergrad Sci/Eng

GAVRIIL TSECHPENAKIS
web: cs.iupui.edu/~gavriil/

PERSONAL STATEMENT

My research domain is Computer Vision and Pattern Recognition. My focus is on handling data uncertainties in classification, modeling, and prediction from image data and beyond. My *hypothesis* is that application of mathematical/computational methods can especially help with ambiguities in the data, outliers and incomplete data, and can ultimately help create new hypotheses and directions in different domains. While keeping my core theoretical background in Computer Vision and Pattern Recognition, the application domains of my interest are robotics, computational neuroscience, as well as biomedical imaging.

Keywords: Appearance- and model-based 2D/3D tracking, structure from motion, structure from temporal feature saliency, unsupervised object detection, graphical model-based structure modeling, point cloud/feature/patch/object-based image registration, multi-modality fusion (features/decision).

EDUCATION

Postdoctoral researcher, Computer Science Nov 2003 -- Dec 2006
Rutgers University, NJ, USA

- Real-time face and gesture tracking in interview settings from monocular videos
- Three-dimensional tracking of hand articulations from monocular videos
- Machine learning-driven deformable models for object segmentation

PhD, Electrical & Computer Engineering Aug 1999 -- Jun 2003
National Technical University of Athens, Greece
Thesis: Moving object detection and tracking from static and mobile cameras

Dipl.-Eng. (5-year Diploma) Aug 1994 -- Jul 1999
National Technical University of Athens, Greece

APPOINTMENTS

Associate Professor (tenured) Jul 2014--present
Indiana University-Purdue University Indianapolis,
USA Dept. of Computer and Information Science

Assistant Professor (tenure-track) Aug 2010--Jun 2014
Indiana University-Purdue University Indianapolis,
USA Dept. of Computer and Information Science

Research Assistant Professor Jul 2008--Jul 2010
University of Miami, FL, USA
Dept. of Computer Science

Researcher, Visualization Program Director Jul 2008--Jul 2010
Center for Computational Science, University of Miami, FL, USA

Visiting Assistant Professor Jan 2007--Jun 2008
University of Miami, FL, USA
Dept. of Electrical and Computer Engineering

AWARDS

2017 US National Research Council (National Academies of Science) Fellowship
2016 Research Frontiers Trailblazer Award, Indiana University-Purdue University Indianapolis 2014 School of Science Research Award, Indiana University-Purdue University Indianapolis 2013 CAREER Award, US National Science Foundation

Honglang Wang

CONTACT INFORMATION
LD 270B, 402 N.Blackford St.,
Department of Mathematical Sciences
Indianapolis, IN, 46202
Phone: (317) 274-7858
E-mail: hlwang@iupui.edu
Homepage:
http://www.math.iupui.edu/~hlwang

RESEARCH INTERESTS
Empirical likelihood method and its applications, statistical analysis for functional and longitudinal data, high dimensional statistical inference and its applications, nonparametric smoothing methods, missing data analysis, bioinformatics and machine learning

EDUCATION

Michigan State University, East Lansing, MI

Ph.D., Statistics, August 2015

Dual Ph.D., Quantitative Biology, August 2015

Zhejiang University, Hangzhou, China

M.S., Mathematics, Jul. 2010

Tianjin University, Tianjin, China

B.S., Mathematics and Applied Mathematics, Jul. 2007

PUBLICATIONS

1. Honglang Wang, Ping-Shou Zhong, Yuehua Cui. "Empirical Likelihood Ratio Tests for Coefficients in High Dimensional Heteroscedastic Linear Models." Statistica Sinica, 2017, accepted
2. Fang Han, Hongkai Ji, Zhicheng Ji, Honglang Wang. "A Provable Smoothing Approach for High Dimensional Generalized Regression with an Application in Genomics." Electronic Journal of Statistics 2017, Vol. 11, No. 2, 4347-4403.
3. Fang Han, Xi Chen, Honglang Wang, Lexin Li, Brain S. Ca↵o. "Robust Graph Change-point Detection with Application to Brain Evolvement Study." Submitted to Journal of the American Statistical Association.
4. Honglang Wang, Ping-Shou Zhong, Yuehua Cui, Yehua Li. "Unified empirical likelihood ratio tests for functional concurrent linear models and the phase transition from sparse to dense functional data." Journal of the Royal Statistical Society: Series B, 2017, accepted.
5. Pratik Nalawade, Kate Ansah-Koi, Khairi Reda, Fang Li, Honglang Wang, Wei Zheng. "The E↵ects of Spatial Frequency and Colormap Characteristics on the Perception of 2D Pseudocolor Scalar Fields." Submitted
6. Xu Liu, Honglang Wang, Yuehua Cui. "Statistical identification of gene-gene interactions triggered by nonlinear environmental modulation." Current Genomics, 2016, 17(5): 388-395.
7. Honglang Wang, Tao He, Cen Wu, Ping-Shou Zhong, Yuehua Cui. "A powerful statistical method identifies novel loci associated with diastolic blood pressure triggered by nonlinear gene-environment interaction." BMC Proceedings 2014, 8(Suppl 1):S61 (17 June 2014)

RESEARCH GRANTS
• (Subrecipient PI with PTE PI: Dr. C. Robin Buell) Unraveling the Heterozygosity, Allelic Composition, and Copy Number Variation of Potato, $14,8122016-2017

Yuni Xia
Associate Professor
Department of Computer and Information Science
Indiana University-Purdue University Indianapolis (IUPUI)
Phone: (317) 274-9738
Web: http://www.cs.iupui.edu/~yxia
E-mail: yxia@cs.iupui.edu

a. Professional Preparation

- Ph.D. in Computer Science, Purdue University, 2005.
- M.S. in Computer Science, Purdue University, 2002.
- B.E. in Computer Science, Huazhong University of Science and Technology, 1996.

b. Appointments

- CIS Department, IUPUI, Indianapolis, IN: Associate Professor, 2012-Present.
- CIS Department, IUPUI, Indianapolis, IN: Assistant Professor, 2005-2012.

c. Awards and Honors

1. Trustee's Teaching Award, IUPUI, 2018
2. Best Demo Award, International Conference on Database System for Advanced Application (DASFAA), 2011
3. Scalable Data Analytics Innovation Award, IBM, 2010
4. Techpoint Mira Award, with Senior Care Navigation System development team at My Health Care Manager LLC, Indiana TechPoint Organization, 2010
5. Research Venture Award, IUPUI, 2009
6. Trustee's Teaching Award, IUPUI, 2009
7. Real Time Innovation Award, IBM, 2008

d. General Summary

Dr. Xia's research is on data mining with a focus on mining data streams and uncertain data. She works on designing new algorithms which incorporates data probabilistic information into the mining process in order to extract more accurate and reliable patterns from the data. She also collaborates with researchers in medicine and healthcare and works on biomedical data mining. She has published more than 60 peer-reviewed publications in journals and conference proceedings. Her current and past research has been funded, as the PI or Co-PI, by the National Science Foundation, IBM, Department of the Army and State of Indiana.

d. Recent Publications

i. Jiaqi Ge, Yuni Xia, Jian Wang, Chandima Nadungodage, Sunil Prabhakar, Sequential Pattern Mining in Databases with Temporal Uncertainty, Journal of Knowledge and Information Systems, 51(3): 821-850, 2017

ii. Chandima Hewa Nadungodage; Yuni Xia; John Jaehwan Lee, GPU Accelerated Online Recommendations for Continuous Data Streams, Journal of Knowledge and Information Systems, 53(3): 637-670 (2017)

iii. Chandima H. Nadungodage, Yuni Xia, Jaehwan Lee, GPU-accelerated Outlier Detection for Continuous Data Streams, the IEEE International Parallel & Distributed Processing Symposium, 2016

iv. Jiaqi Ge, Yuni Xia, Distributed Sequential Pattern Mining in Large Scale Uncertain Databases, The Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD) 2016