Welcome
Data-Driven Discovery

Teik C. Lim, PhD

Provost and Vice President for Academic Affairs
Completed/Underway (select cases):
- New degrees/certs in data science & analytics in CoB, CoE and CoS to meet workforce needs
- DataCAVE – data analytics lab in Library, for software access, collaboration and training needs
- Mavs Dataverse – repository for citable research data, to increase reach/impact of UTA research
- Big Data Analytics Research Center housed in CoE

Planning (select items):
- Cybersecurity in human and societal dimension
- Data Science Clinic
- Industry-university collaboration in SMART megacity analytics
- Interdisciplinary thrust in digital humanities
- Artificial intelligence and machine learning
POSSIBLE NEW/ENHANCED ACADEMIC PROGRAMS OR UNITS

COS
Data Science and Analytics
Bioinformatics

CONHI
Health Care Informatics
Healthcare simulation
Bioinformatics
Biotech

CAPPA
SMART Cities

COLA
Digital Humanities
Media/Digital Comm.

COB
Business Analytics

COE
Data Science & Analytics
SMART Cities
Bioinformatics
Cybersecurity
The CoB is positioning itself to be a leader in Business Analytics and is committed to:

- Preparing students for a successful career in business analytics
- Pursuing research excellence in data driven discovery
- Engaging industry partners
Graduate programs:

- MS Business Analytics
- MS Economic Data Analytics
- MS in Marketing Research

Undergraduate Programs:

- BS Business Analytics (Fall ‘19)
- Undergraduate minors in various business disciplines
Offers technical and business perspective of data

“Hands on” experience for students, allowing them to apply advanced technical skills and solving real world business problems

Excellent growth prospects

Recent student placements - Amazon, Capsule8, Deloitte, GM Financial, PwC, Pier 1 Imports
MS Economic Data Analytics

- Provides students with quantitative skills that allow them to formulate pricing strategies, cost management, process flow, demand forecasting and customer acquisition valuation

- Recent student placements - Cottonwood Financial, Mary Kay, GM Financial, 2M Research, Targetbase, Buxton
Hands-on program designed to prepare students for careers in marketing research

Students meld logic with creativity, quantitative data with qualitative insights, and intelligence to solve marketing problems and create business opportunities

Recent student internships - Enterprise, Northwestern Mutual, Hilti, State Farm, Vivint and Sellmark
Undergraduate Marketing students analyzed customer feedback data and recommended a new marketing strategy for a local popcorn business.
Engaging the Business Community

4th Annual Analytics Symposium

- Held on March 29th 2019, attracted 138 attendees, including 80 professionals from 30 companies
- In conjunction with the symposium, 38 student teams from UTA, UTD, UNT, and SMU participated in an industry sponsored Student Analytics Competition
Executive Certificate in Business Analytics

- Custom Certificate Program developed in cooperation with a Global Fortune 500 company

- Designed to reskill executives and promote a data driven decision making culture
Sample of Research topics:

- Twitter data analysis - understanding health related lifestyle behavior
- Performance of crowd funding projects
- Using deep learning to predict supply chain performance
- Understanding the human brain engaged in economic decision making
Sample of Cooperative Data Driven Discovery between Colleges:

- At UTA: Several MS BA courses are taken by CSE students as electives, COB faculty collaborate with other college faculty on various projects.
- Researchers from other colleges can use the Bloomberg terminals in the new Financial Market Labs.
- Researchers can use the financial markets databases that COB/UTA subscribe to such as Compustat, CRSP, and Option Metrics.
THE UNIVERSITY OF TEXAS AT ARLINGTON
DATA-DRIVEN DISCOVERY SYMPOSIUM
Big Data in Education

Teresa Taber Doughty, PhD
Dean, College of Education
What are the characteristics of Big Data?

- **Velocity**: How fast data are changing
- **Volume**: Amount of available data
- **Variety**: Different kinds and sources of data

Gartner, 2012
What does Big Data mean in Education?

Meaningful Insight
- Students
- Instructors
- Programs
- Institutions

Real-Time Data
- Dashboards
- Statistical Analysis
- Machine Learning
- Data Modeling
- Data Mining

Transforming Education
- Grades
- Outcomes
- Evidence
- Continuous Learning

Personalized Learning

From: Desire2Learn
Higher Education

Student Retention
- First year attrition rates exceed 25%
- Some states reach 40%
- Only 1 in 2 students ever complete a degree

Time to Degree
- 60% FT students complete 4-yr Bachelor’s within 8 years
- 24% PT students complete bachelor’s in 8 years
- 20% take more courses than needed

Degree Completion
- 75% students are non-traditional
- 40% students are not academically prepared
- 40% are part time

From: Complete College America, *Time is the Enemy Summary*
How Are Big Data Sets used in Education?

- personalized instruction
- responsive formative assessment
- actively engaged pedagogy
- collaborative learning
Questions we might answer in Education using Big Data:

• How effective are standardized annual K-12 state assessments in predicting future outcomes (graduation rates, post-school employment, success in STEM fields, etc…)?

• What are the employment rates of individuals with disabilities who complete high school? A 4-year degree? Graduate degree? How does type of disability impact employment rates?

• Is there a time-to-degree difference for students who complete a traditional, F2F program versus a blended or online post-secondary program?
Current Projects in the COEd

- Dr. Bradley Davis
  - Examining big data set related to longitudinal data relating to Texas public schools
  - [https://www.utdallas.edu/research/tsp-erc/data-holdings.html](https://www.utdallas.edu/research/tsp-erc/data-holdings.html)

Figure 2:

Alluvial Diagram Depicting the Career Pathways of Texas Educators with Principal Certification

From: Davis, B.W., & Anderson, E. *What becomes of leadership preparation program graduates? Investigating changes in role assignment over the years following credentialing.* Manuscript in progress.

Note: This image can be viewed in full resolution here: [http://i.imgur.com/8GAUwzY.jpg](http://i.imgur.com/8GAUwzY.jpg)
• Dr. Maria Trache
  – Restricted and unrestricted data sets: National Center of Education Statistics under an IES/NCES license and NSF
  – Answers questions about gender differences in labor market outcomes; employment and earnings of international STEM graduates of U.S. universities; transfer student success
Current Projects in the COEd

• Dr. Jodi Tommerdahl
  – Visual Accent Trainer
  – Used large language samples to look for “statistically normal language” to better identify abnormal language and language disorders
  – Predictive modeling (subcategory of learning analytics) within a virtual educational system that may impact educational policy and instructional strategies
Potential Collaborations

- Cost/Benefit Analysis
- School violence
- Relationship to healthy communities
- Link between zip code and education outcomes

Education
Big Data Research - Opportunities and Challenges

Gautam Das, PhD

Professor, Department of Computer Science & Engineering
Big Data - Opportunities

- Increased availability of data yield accurate analysis in health care to genomics, business to physics
- Insights from big data leads to confident decision making
- Better decision leads to greater efficiencies, cost reductions, and reduced risks

Data Driven Discovery: The exponential growth & availability of big data presents opportunities to fuel decisions at every level of society
Big Data - Challenges

- High Volume
- Fast Changing
- Complex
- Heterogeneous

*Challenges in data collection, cleaning, storage, management, analysis*
The Big Data Science Pipeline

Data Collection

Cleaning & Storage

Management & Query

Analysis & Applications
Core Big Data Research Opportunities

Big Data Storage and Management

Challenges in storing, managing & querying big data

Data collection technologies such as sensors & cyber-physical systems (e.g., IoT)

High-performance computing systems that rely on fast networks & distributed and multi-core systems

Cloud computing platforms
Big Data Storage and Management

Data collection technologies - sensors & cyber-physical systems (e.g., IoT)

Chris Ding  
CSE

Yan Wan  
EEE

Chengkai Li  
CSE

Ming Li  
CSE
Big Data Storage and Management

High-performance computing systems that rely on fast networks & distributed and multi-core systems

Hong Jiang
CSE

Ishfaq Ahmad
CSE

Jia Rao
CSE
Big Data Storage and Management

Cloud computing platforms

Mohammad Islam  
CSE

William Beksi  
CSE

Jia Rao  
CSE
Core Big Data Research Opportunities

**Advances in Big Data Mining and Machine Learning**

Scaling of existing data mining and machine learning algorithms in big data environments (E.g., scalable clustering and classification)

Develop new machine learning algorithms to take advantage of advances in hardware, sophisticated algorithms and scalable software framework (E.g., deep learning)
Big Data Mining and Machine Learning

Scaling existing data mining and machine learning algorithms in big data environments

- Sharma Chakravarthy, CSE
- Leonidas Fegaras, CSE
- Song Jiang, CSE
- Hao Che, CSE
- Ishfaq Ahmad, CSE
- Mohammad Islam, CSE
- Jia Rao, CSE
Big Data Mining and Machine Learning

Develop machine learning algorithms to take advantage of advances in hardware and scalable software framework
Big Data Applications Research

- Computer vision
- Natural language processing
- Autonomous cars
- Robotics
- Genomics
- Law
- Business
- Healthcare
- Engineering
- Physical and Social Sciences
Big Data Applications Research

Engineering

Andrew Makeev
MAE

Yan Wan
EEE

Stephen Mattingly
Civil Eng

Shouyi Wang
Industrial Eng

Seyedali Abolmaali
Civil Eng

Anand Puppala
Civil Eng

Jay Rosenberger
Industrial Eng

Kate Hyun
Civil Eng
Big Data Applications Research

Healthcare

Gautam Das
CSE

Shouyi Wang
Industrial Eng

Won Hwa Kim
CSE

Ishfaq Ahmad
CSE

Dajiang Zhu
CSE

Chris Ding
CSE

Jean Gao
CSE

Junzhou Huang
CSE

Fillia Makedon
CSE

Vassilis Athitsos
CSE
Big Data Applications Research

Robotics
- Fillia Makedon
- William Beksi

Computer Vision
- Andrew Makeev
- Vassilis Athitsos
- Won Hwa Kim
- William Beksi

Business
- Mohammad Islam
- Ming Li
- Sridhar Nerur

Physics and Social Sciences
- Fillia Makedon
- Yan Wan
- Kaushik De
Big Data Applications Research

NLP
- Chengkai Li
  CSE
- Deokgun Park
  CSE

Security
- Shirin Nilizadeh
  CSE
- Jiang Ming
  CSE
- Jeff Lei
  CSE
- Ming Li
  CSE

Genomics
- Mohammad Islam
  CSE
- Jay Rosenberger
  Industrial Eng

Jean Gao
CSE
Proposal for BigDAC: Big Data Analytics (BDA) Center

Director: Gautam Das

Mission

- Develop “grand-challenge” research projects in BDA
- Deliver world-class research via papers, patents, and products
- Serve as a hub for UTA faculty, researchers, students as well as external collaborators to engage on BDA issues
- Reach out to industry, government, and nonprofits to understand their BDA needs, and partner with them to develop new research projects and educational programs
- Create innovative BDA technologies and help to transfer such technologies into the real world by encouraging entrepreneurship and partnerships with industry and organizations

Current Industry Collaborations

Members and Collaborators

- 15 UTA faculty, 11 external researchers including several industrial collaborators

Research Focus

- Big Data Storage and Management
- Big Data Mining and Machine Learning
- Big Data Analytics Applications
THE UNIVERSITY OF TEXAS AT ARLINGTON
DATA-DRIVEN DISCOVERY SYMPOSIUM
Data-Driven Physics Discovery

Kaushik De, PhD

Professor, Department of Physics and Director, High-Energy Physics Center of Excellence
Introduction

• UTA High Energy Physics at the forefront of data driven discoveries for 20 years
  – Scientific data drives scale and complexity
  – Computational advances drive discoveries
  – Infrastructure and software equally important
• A look at the past, present, and future
The ATLAS Experiment

 Worlds largest experimental apparatus located 100m underground at the Large Hadron Collider (LHC)

LHC is a high energy proton smasher near lake Geneva, in a 27 km long tunnel snaking under France and Switzerland

UTA in ATLAS since 1995
Why build ATLAS?

The Nobel Prize in Physics 2013

Study the fundamental properties of our universe

Particles and forces

Around 750 journal publications

Continuing study of Higgs; search for dark matter, study Standard Model … till >2035
• The largest “camera” ever built
• And the fastest – few hundred million pictures every second
• Each picture contains ten million bits of complex information
• Scientific data driven discoveries – scale and complexity

Parts were built by students at UTA
ATLAS Computing Ecosystem

Worldwide LHC Computing Grid (WLCG)

- Computing centers are distributed worldwide
- ~150 clusters
- Heterogeneous
- Independently maintained
- Provides ~300k CPU cores on average, one million burst
- ~350 PB storage

Dozens of applications
Dozens of workflows
Worldwide community of thousands of users
Infrastructure and Software used globally
Driven by innovations
Inventing PanDA

– PanDA software was started by UTA and BNL a dozen years ago
– Allows us to use computers at data centers around the world to solve data driven science problems
– Pre-cursor to cloud computing
– Cutting edge of US innovations
PanDA Scale – Jobs Completed

Completed jobs
365 weeks from Week 26 of 2009 to Week 26 of 2016

Total Per Month

Maximum: 38,501,894, Minimum: 0.00, Average: 23,811,827, Current: 30,491,396
• Truly worldwide data movement, storage and processing
• And massive simulations to understand the data
UTA Leading the Way

- LHC computing center at UTA
  - ~12,000 CPU cores
  - ~7,000 TB disk space
  - Global computing center for ATLAS
  - PanDA, AI, deep learning…

Future challenges – LHC data growing by factors of 5-10 in the next decade. Need new innovations.
THE UNIVERSITY OF TEXAS AT ARLINGTON
DATA-DRIVEN DISCOVERY SYMPOSIUM
Registrant Speakers

- Ishfaq Ahmad
- David Arditi
- William Beksi
- Sharma Chakravarthy
- Morgan Chivers
- Rebekah Chojnacki
- Muhammad Huda
- Ashley Lemke
- Hanli Liu
- Peace Ossom Williamson
- Kenneth Roemer
- Leili Shahriyari
- Jianzhong Su
- Charles Travis
- Ahoura Zandiatashbar