Forming Interdisciplinary Partnerships

Auroop Ganguly, Northeastern
Srinivas Aluru, Georgia Tech
Overarching Themes

• “Methods” Represented
  • Applied Math
  • Statistics
  • Physics
  • Algorithms
  • Optimization
  • Networks
  • Machine Learning
  • Data Mining
  • Parallel & Distributed Computing
  • High Performance Computing
  • Visualization
  • …

• “Domains” Represented
  • Bioinformatics: Biology
  • Bioinformatics: Genomics
  • Bioinformatics: Invasive Species
  • Bioinformatics: Public Health
  • Geoscience: Climate
  • Geoscience: Weather
  • Geoscience: Hydrology
  • Engineering: Infrastructure
  • Engineering: Transportation
  • Engineering: Aerospace
  • Engineering: Environmental
  • Engineering: Electrical
  • …
Overarching Themes

- **Step 1**: Data science “solutions looking for problems”
  - A lot of focus on Methods and not enough focus on domain Discoveries
- **Step 2**: “Domain applicable” solutions
  - Collaboration with domain expert and Continuous feedback
- **Step 3**: Generalizable data science solutions
  - Methods or approaches that may generalize across a few domains
- **Step 4**: Application logic or “boundary objects”
  - Adaptations of Big Data methods to application modules which can be generalized
- **Step 5**: Learning adaptations of application logic
  - Solutions that can generalize across multiple domains
- **Step 6**: Focus on data-driven “Discovery” across domains
  - However, validation remains a challenge
- **Step 7**: Use-based and modular validation
  - Cross thematic workshops, ontology / taxonomy, and knowledge repository
- **Step 8**: Domain Questions → Problems → Taxonomy of Methodological Needs
  Big Data Methods → Solutions → Taxonomy of Available Methods

Mapping the Square Peg to the Round Hole
Recent Successes

• Human Genome Analysis
• Climate & Weather modeling
• Self-driving Cars
• The marriage of computing and social sciences
• Brain Initiative
• Precision Medicine
Major Obstacles

• “Lost in Translation” (Methods $\leftrightarrow$ Domains)
  • From Methods Solutions to Domain Problems
  • From Domain Questions to Methods & Tools
  • No emphases or incentives to gradually master both

• No Inter-Domain Communication
  • No Cross-Domain Knowledge Transfer on Big Data
  • Domains need to understand each others’ nuances
  • Cross-domain lessons from successes and failures
  • Side benefits of such cross-domain fertilizations immense

• Lack of cross-agency data/problem sharing
  • Inadequate data sharing
  • Inadequate incentives for PIs to solve problems jointly
Areas of Neglect

• “Lost in Translation” (Methods ↔ Domains)
  • No formal incentives for the crucial communication phase
  • No incentives for early career PIs to develop cross-discipline (Data + Domain) or cross-domain expertise

• No Inter-Domain Communication
  • No repository or benchmarks or taxonomy for cross-domain learning of Big Data applications and success/failure stories with domain nuances
  • No publication venues considered high impact enough for cross-domain lessons learned or for interdisciplinary work

• Lack of cross-agency data/problem sharing
  • Lack of incentives and many barriers to inter-agency calls
  • No easy access to multiple agency data or problems
Strategic Priorities & Investments

• “Lost in Translation” (Methods ↔ Domains)
  • Fund regular workshops run by PIs, one (or more) from methods and one (or more) from domains
  • Fund “seedling” (not just “eager”) proposals just to encourage communications with outcomes such as position papers and full interdisciplinary proposal development

• No Inter-Domain Communication
  • Develop new solicitations for funding the development of cross-domain benchmarks and repositories
  • Develop new solicitations that require 3-way (or >3-way) cross-pollination such as Big Data – Public Health; Big Data – Climate; and Climate – Public Health with an integrating thread

• Lack of cross-agency data/problem sharing
  • Develop cross-agency data sharing and incentivize their use