FEW Nexus: Informatics Challenges

Vasant G Honavar
Professor of Information Sciences and Technology
Director, Center for Big Data Analytics and Discovery Informatics
vhonavar@ist.psu.edu
http://faculty.ist.psu.edu/vhonavar/

The Pennsylvania State University
Food-energy-water nexus

- Complex interactions between energy, water, food
  - Land for food (agriculture)
  - Land for water (water capture, reservoirs)
  - Land for energy (solar and wind farms)
  - Water for food (agriculture)
  - Water for energy (hydropower)
  - Energy for land (built environment)
  - Energy for water (water supply)
  - .......
- Mediated by
  - climate change (increased frequency of extreme weather events, changes in regional climate and hydrologic cycles)
  - population dynamics (changes in population, migration patterns, demographics)
  - human influences on land use, climate, etc.
  - policy landscape (water, energy, agriculture...)

Vasant Honavar, NSF Workshop on Data Science Challenges in FEW Nexus, Oct 5, 2015
Vasant Honavar, NSF Workshop on Data Science Challenges in FEW Nexus, Oct 5, 2015
FEW Nexus in Context


Vasant Honavar, NSF Workshop on Data Science Challenges in FEW Nexus, Oct 5, 2015
FEWS presents a big data problem

- Data volume, variety, velocity, variety, veracity ...
- Multiple modalities
- Multiple data types
- Multiple spatial and temporal scales
- Multiple levels of granularity
- Varying levels of data and model uncertainty
- Varying data and model quality
FEW Nexus presents a big data problem

Global Water Stress Map (Water Stress = Withdrawals/Available Flow)
Source: World Resources Institute Aqueduct Water Risk Atlas
Taken from DOE June 2014 report

Vasant Honavar, NSF Workshop on Data Science Challenges in FEW Nexus, Oct 5, 2015
Informatics challenges: Data Management

- Representing and reasoning about data
  - Metadata and ontologies
  - Sensing process
  - Uncertainty
  - Provenance
- Integrating data
  - Of multiple types (including omics, measurements, imagery, reports, social media)
  - Across multiple modalities
  - Spatial and temporal resolution
  - Of varying quality, uncertainty
- Managing data
  - Storage, indexing, archiving
  - Query languages, interfaces
- Visualization
Informatics challenges: Data Analytics and Modeling

- Analyses, models, and predictions that
  - Are comprehensible, accurate, communicable
  - Are testable
  - Are consistent with, and take advantage of physics
  - Incorporate data from social, demographic, and other types of data
  - Close the language gap between model builders and model users
  - Provide causal explanations
  - Reconcile explanations at multiple levels of abstraction
  - Support reasoning about alternative scenarios and stakeholder priorities and preferences at decision relevant scales

Vasant Honavar, NSF Workshop on Data Science Challenges in FEW Nexus, Oct 5, 2015
Informatics Challenge: Data Analytics and Modeling

- Machine learning methods focus on constructing predictive models typically from observational data
- Use of off-the-shelf machine learning algorithms leads to models that
  - Can be hard to interpret
  - Can Ignore, or worse, violate physics
  - Introduce a language gap between model builders and model users
Informatics Challenge: Data Analytics and Modeling

- New methods needed for
  - Building models from spatial and temporal (longitudinal) data
  - Multi-scale, multi-view modeling
  - Data assimilation and model adaptation
  - Causal inference (especially for complex spatio-temporal data), causal transportability
  - Scalable hierarchical Bayesian methods
  - Physics-based machine learning
Informatics Challenge: Data Analytics and Modeling

• New methods needed for
  • Spatial (as opposed to topological) network modeling
  • Circumventing the need for sensitive data (e.g., precise location, land use)
• Literature based discovery – connecting the dots across disparate scientific disciplines
• Closing the language gap between model builders and model users
• Closing the data – model – hypotheses - prediction – observation-experimentation loop

Vasant Honavar, NSF Workshop on Data Science Challenges in FEW Nexus, Oct 5, 2015
Informatics challenges: Model assessment

- Need effective approaches to quantifying and communicating
  - Model performance
  - Model utility
  - Uncertainty associated with model structure, parameters, and predictions
  - Model robustness
Informatics challenges: Decision support

- Need effective approaches to multi-stakeholder
  - Preference and tradeoff elicitation
  - Negotiation
  - Decision making
- Visualizing data, models, decisions, impacts
Informatics challenges: Community Infrastructure

• Engaging a broad and diverse community to contribute requires
  • Curated data sets for developing and comparing models
  • Interoperability of software and platforms
  • Community standards for sharing data and models
  • Shareable workflows for reproducible analyses
  • Incentive structures to foster collaboration
• Training of a new generation of scientists and practitioners with deep expertise in informatics and one or more areas of FEW nexus
Thank you!