

Concluding Thoughts

Tariq Samad

Honeywell Automation and Control Solutions

Minneapolis, MN, U.S.A.

tariq.samad@honeywell.com

IMA Workshop on “Data-Driven Control and Optimization,” Dec. 4-6, 2002



Three Great Things About the Twin Cities

- The weather's bracing in early December
- Restaurants are pretty good (at any time of year)
- The Institute for Mathematics and its Applications
 - an international jewel

Thanks Scot, Doug, Fadil, Judie!

- Model-predictive control in the process industries
 - successful applications in 1970s, without theoretical underpinnings,
 - theory developed later
 - practitioners led the way, researchers followed
- The Matlab *System Identification Toolbox*
 - developed by Prof. Lennart Ljung, 1980s
 - rigorous analytical framework, then a software package
 - widely adopted by industry

Basic theory usually the preserve of academia, new applications of industry, but the spark of innovation can arise anywhere

- Too many criteria, and they're application dependent
 - prediction accuracy on new data
 - false positive vs false negative errors (and function approximation analogs)
 - model complexity (usually, but not always, lack thereof)
 - understandable by users
 - cost of model development
 - cost of model maintenance
- No technique can be expected to be globally superior
- No theory can be expected to be this comprehensive

- Skill level of human user
 - e.g., they don't teach reproducing kernel Hilbert spaces in high school—or medical school
- Problem dimensionality and sample size
 - DNA microarrays versus power grids versus batch reactors
- Regulatory requirements
 - FDA, FAA, EPA, FERC, ...
- Cost of upgrading system
- Response time requirements
- Stationarity of system
- Accuracy requirements

What improvement in accuracy will overcome need to train users, obtain regulatory approval, upgrade equipment?

- Intelligence and security
- Power systems
- Process industries
- Veterinary treatment
- Genomics
- Systems biology
- District heating networks
- Financial markets
- Probabilistic mapping
- Chronic relapsing disorders
- Supply chain management

It would be nice if “data-driven” implied “domain knowledge not needed” . . .

. . . but real and practically important domains have particular characteristics

- power systems and markets
- microarray data mining
- batch bioreactor

Must every researcher become an expert in some complex domain?

Theory can be developed in isolation, but deriving real-world benefit requires multidisciplinary/multifunctional collaboration

The Long-Term View

First “artificial neural network” model: McCullough and Pitts, 1940s

The first wave of hype: Rosenblatt Perceptron, early 1960s

From oversell to overkill: Minsky and Paper, mid-1960s

The new wave begins: Hopfield, McLelland & Rumelhart, ...: early 1980s

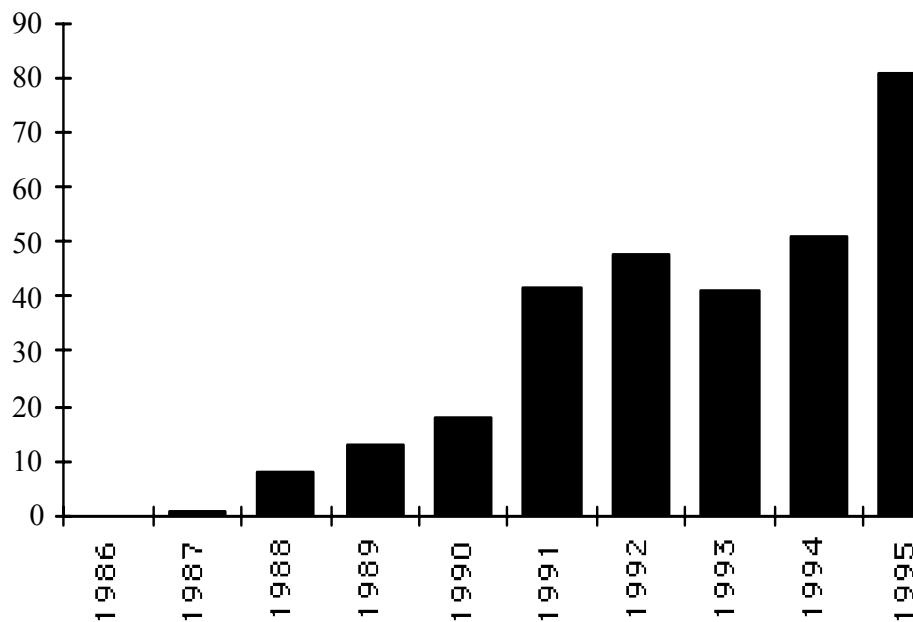
Second wave of hype...

Control and statistics communities get involved

- heuristics lead to rigor

Successful practical applications

Current status: Not a revolutionary technology, but a useful tool in the data analyst’s toolbox



Number of papers with “neural networks” or synonyms in title presented at the American Control Conferences, 1986-1995



- Data repositories in existence, widely used by research community
- Need industrial-strength data sets
 - real problems, sanitized data
 - substantial potential benefit to contributing organization—only they can exploit “better” modeling approaches for the sanitized problem
 - possible role for academic centers, consortia, or not-for-profit institutions (e.g., EPRI)

- We lack standard formulations for data-driven modeling
- Separate formulation from method
- Facilitate comparisons, evaluation, and application

- $y = f(x; \theta)$ is just one kind of data-driven model ...
- Influence diagrams and limited memory influence diagrams (LIMIDs)
 - modeling workflow, and associated information flow
 - efficiently maximizing utility in multistage decision procedure
 - what information should be taken into account in making a decision?
 - cooperative decision making
 - close connection to heuristic dynamic programming and related methods, but explicit multi-stage decision-making perspective is illuminating

- Systems biology models
 - action and interaction of cells within the organism
 - action and interaction of genes within cell
 - integrating description levels
 - ontologies
 - models being developed manually, but data is becoming available

- Dynamic dynamical systems
 - dynamical systems that undergo structural changes—components and connectivity
 - .NET, critical infrastructures, autonomic computing, C4ISR
 - How do we realize just-in-time plug and play architectures?
 - How do we design control strategies?

- Model-predictive control
- System optimization
- Route planning
- Cautious switching control
- Optimized multistage decision making
- “Gaming” the system
- System security

Stochastic and probabilistic aspects are central to data-driven control and optimization

- Regularization and model complexity measures
- Dow's stacked analytic neural nets
- Pruned information flow
- Suboptimal distributed optimization
- Offline/online partitioning of computation
- Parametric approximation, information geometry, relevance-based weighting, exploiting correlation structure

- The same approximation may facilitate the solution of one problem and render the solution of another, seemingly similar one, intractable

- Workshop is unique in some ways
 - theory and practice
 - broad view of data-centricity
 - multiple application areas
- Published proceedings may be a useful resource for multiple communities
 - Springer-Verlag bound volume
 - broaden title to “Data-Centric Modeling, Control, and Optimization”?
- Organizers will solicit additional chapters
 - suggestions for invitees welcome
- Suggested timeline
 - invitations to additional desired contributors: 31 December 2002
 - preliminary table of contents, 31 January 2003
 - chapters due 30 April 2003 (send to IMA)
 - material to Springer: 1 June 2003
 - book out: 1 August 2003
- IMA will provide LaTeX macros, do final typesetting
 - suggested length: 12-20 pages in IMA format