Data-Driven Optimization

John Turner
Assistant Professor
The Paul Merage School of Business
University of California, Irvine

October 24, 2014: UCI Data Science Initiative
Analytics – Unlocking the Potential of Big Data

➡️ Descriptive – “What is happening?”
Querying, Reporting, Data Capturing, Filtering & Analysis

➡️ Predictive – “What will likely happen?”
Statistical Methods (Regression), Forecasting & Data Mining

➡️ Prescriptive – “What should we do?”
Optimization, Simulation, Quantitative Models
Optimization

\[ \begin{align*}
\text{min } & \quad f(x) @ \\
\text{s.t. } & \quad g_{\downarrow i}(x) \leq 0, \\
& \quad h_{\downarrow i}(x) = 0, \\
& \quad x \in X
\end{align*} \]
Applications for Optimization

- Chip design
- Logistics
- Supply chain management
- Network design
- Scheduling
- Production planning
- Timetabling
- Marketing campaign design
- Portfolio optimization
- Air traffic routing
Problems, Formulations, and Instances

1. Define Problem
2. Construct Mathematical Model (Formulation)
3. Solve one or more Instances
4. Communicate Results / Automate the Process

Algebraic Manipulation

Solver (CPLEX, Gurobi, KNITRO, etc.)
Computational Challenges

\[ f(x) \] @  
...m@  
...p@  \[ x \in X \]

- Nonconvex objective
- Too many constraints
  \[ s.t. \quad g \downarrow i(x) \leq 0, \]
  \[ h \downarrow i(x) = 0, \]
- Too many variables
- Nonlinear constraints
- Nonconvex domain (e.g., discrete \( X \))

\[ \Rightarrow \text{Large-Scale Instances Require Specialized Decomposition or Approximation Schemes} \]
Two Examples

1. Planning Online Advertising

2. Locating Trauma Centers at Nation-Scale
Different “viewer types”
Different “viewer types”
Different “targeting requirements”
Impression goals for each advertiser
We use optimization to find an “optimal allocation”
Combinatorial Explosion

Targeting can specify:

- gender
- age
- geographic region
- income level
- day of week / time of day
- the sports you like, etc.

Trillions of possible combinations!!
Supply highly variable
Expected supply close to zero!

Questions:
1. What is “best” way to aggregate?
2. How to interpret aggregate solution?
3. How good is aggregate solution?
Implications

• Allow the optimization to segment viewers into buckets
• This becomes an iterative process
  • Solve → Learn → Revise → Repeat
• Data requirements change at each iteration
Data-Driven Optimization

**Sequential Process**
1. Construct Estimates of Model Parameters (Forecasting / Data Mining)
2. Solve Optimization Model

**Iterative Process**
1. Start with Coarse Approximations for Model Parameters (Cheap Forecasts)
2. Solve Optimization Model
3. Learn & Revise Model
4. Re-sample Data to Improve Forecasts
5. Repeat Steps 2-4

**Take-away:** Can often get to within 1% of optimality with less than 100 judiciously-chosen audience segments
Example 2: Demand for Trauma Care in Korea

- Population ~ 50 million
- Area 100,210 km² (~ Indiana)
- 7 metro cities & 9 provinces
- Number of trauma cases**
  = 190,196*
  ~ 382.1 trauma / 100k

* Stats in year 2008
** Patients with EMR-ISS score ≥ 15
Candidate Sites

Candidate Trauma Center Sites (n=38)

Candidate Heliport Sites (n=54 total: 16 NEMA* + 38 @ TC’s)

*NEMA = Korean National Emergency Management Agency sites (heliports currently used for fire & other EMS missions)
For each approach (decoupled, no-congestion, and integrated),

- indicates the sites chosen for T10H5;
- indicates the sites chosen for T10H15 but not for T10H5;
- indicates the sites chosen for T10H25 but not for T10H15

**Take-away:** Solutions of varying quality can look visually similar.
Successful Patients (%)

Take-away: An integrated model solved approximately is often better than solving two more-precise models in sequence.

---

### 10 trauma centers

- **Integrated**
- **NoCongestion**
- **Decoupled**

### 12 trauma centers

### 14 trauma centers

---

# Helicopters

---

21
Big Data, Big Challenges

Conclusions:

- Prediction and Optimization work well together
- Predict AND Optimize, instead of Predict THEN Optimize.
Data-Driven Optimization

John Turner
Assistant Professor
The Paul Merage School of Business
University of California, Irvine

October 24, 2014: UCI Data Science Initiative