Data-Driven Decision-Making
In Enterprise Applications

Introduction

Rainer Schlosser, Martin Boissier, Matthias Uflacker

Hasso Plattner Institute (EPIC)

April 18, 2019
The World is Full of Decision Problems
What Constitutes a Decision Problem?

Decisions

Objectives

Constraints
How to Approach Decision Problems?

**Decisions** $x$  
When can I do what?  
Identify.

**Objective** $F(x)$  
What do I want to optimize?  
Define.

**Constraints** $C(x)$  
What has to be satisfied?  
Determine.

Data-Driven Decision-Making in Enterprise Applications - Introduction
How to Approach Decision Problems?

Decisions $x$
- When can I do what? Identify.

Impact of $x$
- What happens if a certain decision is made? Estimate.

Objective $F(x)$
- What do I want to optimize? Define.

Constraints $C(x)$
- What has to be satisfied? Determine.

Optimization
- $\max F(x)$ over $x$ such that $C(x)$ is satisfied. Solve!

*Data-Driven Decision-Making in Enterprise Applications - Introduction*
Agenda

- Introduction ✓

- Personal Background

- Goals of the Course & Grading

- Examples: Decision Problems in Data-Driven Applications
Personal Background

- Ph.D. Operations Research (2014), Humboldt-University of Berlin
- Hasso Plattner Institute, EPIC, since 2015
- Field of Research
  - Data-driven decision support
  - Focus on stochastic dynamic models
- Current Areas of Applications
  - Operations management (e.g., dynamic pricing, ordering, advertising)
  - Database configuration (e.g., data placement problems, index selection)
Agenda

- Introduction  ✓
- Personal background  ✓
- Goals of the Course & Grading
- Examples: Decision Problems in Data-Driven Applications
Technical Information

- Credits? 4 SWS (V/Ü), 6 ECTS (graded)
- When? Monday 13.30 - 15.00 / Thursday 11.00 – 12.30
  Start: April 18, 2019,  End: July 11, 2019
- Where? Room D-E. 9/10
- Who? Rainer Schlosser,  rainer.schlosser@hpi.de
  Martin Boissier,  martin.boissier@hpi.de
- Slides? EPIC, Teaching, Summer 2019
Structure of the Course

- **April/May:** Lectures on „Optimization Techniques“:
  (i) Linear Programming
  (ii) Integer Linear Programming
  (iii) Linear/Logistic Regression
  (iv) Dynamic Programming
  (v) Robust Optimization

- **June/July:** Choose Projects, Apply/Extend Suitable Techniques, Work in Teams, Input/Support, Presentations

- **Aug/Sep:** Documentation of Projects Results
## Overview

<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
<th>Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 April 25</td>
<td>Linear Programming</td>
<td>Data-Driven Decision-Making in Enterprise Applications - Introduction</td>
</tr>
<tr>
<td>3 April 29</td>
<td>Integer Linear Programming</td>
<td></td>
</tr>
<tr>
<td>4 May 2</td>
<td>Linear/Logistic Regression</td>
<td></td>
</tr>
<tr>
<td>5 May 6</td>
<td>Exercise Implementations</td>
<td></td>
</tr>
<tr>
<td>6 May 16</td>
<td>Dynamic Programming I</td>
<td></td>
</tr>
<tr>
<td>7 May 20</td>
<td>Dynamic Programming II</td>
<td></td>
</tr>
<tr>
<td>8 May 23</td>
<td>Response Strategies / Game Theory</td>
<td></td>
</tr>
<tr>
<td>9 May 27</td>
<td>Project Assignments</td>
<td></td>
</tr>
<tr>
<td>10 June 3</td>
<td>Robust Optimization</td>
<td></td>
</tr>
<tr>
<td>11 June 13</td>
<td>Workshop / Group Meetings</td>
<td></td>
</tr>
<tr>
<td>12 June 20</td>
<td>Presentations (First Results)</td>
<td></td>
</tr>
<tr>
<td>13/14 June 24/27</td>
<td>Workshop / Group Meetings</td>
<td></td>
</tr>
<tr>
<td>15/16 July 1/4</td>
<td>Workshop / Group Meetings</td>
<td></td>
</tr>
<tr>
<td>17 July 11</td>
<td>Presentations (Final Results), Feedback, Documentation (Aug 31)</td>
<td></td>
</tr>
</tbody>
</table>
Goals of the Course & Grading

- **Goal:** Develop models to compute optimized decisions for data-driven applications
- **Learn:** Optimization techniques
- **Do:** Apply & extend different optimization approaches
- **Grading:**
  - 10% Regular attendance / Personal engagement
  - 20% Results / Homework
  - 30% Presentations
  - 40% Documentation / Paper (End of semester)
Prerequisites

- Programming
  - Parameters, Data Preparation
  - Loops, Recursions, Simulations

- Basic Mathematical Background
  - Sets, Vectors
  - Probabilities, Random Variables, Expected Values

- More does not harm
  - Regression Analysis
  - Experience with Solvers
  - Game Theory
Agenda

- Introduction ✓

- Personal Background ✓

- Goals of the Course & Grading ✓

- Examples: Decision Problems in Data-Driven Applications
Problem Example 1 – Dynamic Pricing

How can we assist an e-commerce merchant in optimizing his/her prices?

*Data-Driven Decision-Making in Enterprise Applications - Introduction*
Impact of Price Decisions and Changing Markets

Characteristics:
- Exits & entries of competitors
- Active and passive competitors
- Price cycles

*price*

![Price Trajectories for ISBN 3980283038](image)

- **Firm 1** (Q: Very Good, R: 97-100%, FC: 80,422)
- **Firm 2** (Q: Good, R: 90-94%, FC: 4,802,970)
- **Firm 3** (Q: Good, R: 90-94%, FC: 56,260)
- **Firm 4** (Q: Very Good, R: 80-89%, FC: 4,802,970)
- **Firm 5** (Q: Very Good, R: 80-89%, FC: 5,156)
- **Firm 6** (Q: Very Good, R: 80-89%, FC: 1,169,441)

*time*
Pricing Options: Price Updates on Amazon

- Price update process on Amazon: (i) request a market situation (ii) optimize price based on demand model, (iii) send price update
Estimation of Price Impacts and Optimization

(1) Estimation of Sales Probabilities

- ca. 10 market situations/day/item with 1-20 firms (100 Mio obs.)
- ca. 2000 sales/month (1 year of data)
- Predict sales probabilities (for time intervals and market situations)

(2) Price Optimization
Estimation of Price Impacts and Optimization

(1) Estimation of Sales Probabilities

- ca. 10 market situations/day/item with 1-20 firms (100 Mio obs.)
- ca. 2,000 sales/month (1 year of data)
- Predict sales probabilities (for time intervals and market situations)

(2) Price Optimization

- Maximize expected discounted long-term profit
- Dynamic programming
Estimation of Price Impacts and Optimization

(1) Estimation of Sales Probabilities

- ca. 10 market situations/day/item with 1-20 firms (100 Mio obs.)
- ca. 2 000 sales/month (1 year of data)
- Predict sales probabilities (for time intervals and market situations)

(2) Price Optimization

\[
\max E(G_t \mid X_t = n, \bar{S}_t = \bar{s}_t), \quad G_t := \sum_{s=t}^{T-1} \delta^{s-t} \cdot \left( (a(X_s, \bar{S}_s) - c) \cdot (X_s - X_{s+1}) - l \cdot X_s \right)
\]

\[
a(n, \bar{s}) = \arg \max_{a \in A} \left\{ \sum_{i \geq 0} \tilde{P}(i, a \mid \bar{s}) \cdot \left( (a-c) \cdot \min(n, i) - n \cdot l + \delta \cdot V\left((n-i)^+, \bar{s}\right) \right) \right\}
\]

\[
V(n, \bar{s}) = \max_{a \in A} \left\{ \sum_{i>0} \tilde{P}(i, a \mid \bar{s}) \cdot \left( (a-c) \cdot \min(n, i) - n \cdot l \right) / \left( 1 - \tilde{P}(0, a \mid \bar{s}) \cdot z \cdot \delta \right) \right\}
\]
Comparison of Performance Results

Comparison: Our *data-driven* strategy vs. the seller’s *rule-based* strategy

<table>
<thead>
<tr>
<th>Strategy</th>
<th>#Books</th>
<th>% Sold (3 months)</th>
<th>Profit per sale (EUR)</th>
<th>Acc. profit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rule-Based</td>
<td>5,534</td>
<td>42 %</td>
<td>2.56 €</td>
<td>100.0 %</td>
</tr>
<tr>
<td>HPI1 (high prices)</td>
<td>5,206</td>
<td>29 %</td>
<td>3.58 €</td>
<td>+40 %</td>
</tr>
<tr>
<td>HPI2</td>
<td>5,407</td>
<td>37 %</td>
<td>3.03 €</td>
<td>+19 %</td>
</tr>
<tr>
<td>HPI3</td>
<td>5,241</td>
<td>44 %</td>
<td>2.94 €</td>
<td>+15 %</td>
</tr>
<tr>
<td>HPI4 (low prices)</td>
<td>5,200</td>
<td>45 %</td>
<td>2.52 €</td>
<td>+6.4 %</td>
</tr>
</tbody>
</table>

KDD 2018

*Data-Driven Decision-Making in Enterprise Applications - Introduction*
Optimal Response Strategies in Duopoly Settings

Question: How do optimal price adjustment strategies look like?

Setting: Infinite horizon, competitor’s response strategy is known

Results:

against $F(a) := \max(a - 1, 1)$
Optimal Response Strategies in Duopoly Settings

Question: How do optimal price adjustment strategies look like?

Setting: Infinite horizon, competitor’s response strategy is known

Results:

\[ a(p) \]

against \( F(a) := \max(a - 1, 1) \) mutual optimal (equilibrium)

Data-Driven Decision-Making in Enterprise Applications - Introduction
Optimal Response Strategies in Duopoly Settings

Question: How do optimal price adjustment strategies look like?

Setting: Infinite horizon, competitor’s response strategy is known

Results: $a(p)$ against $F(a) := \max(a - 1, 1)$ mutual optimal (equilibrium)
Interaction of Self-Adapting Strategies (Short-Term)

- Now, price responses *have to be learned!*
- Both players update their strategies
- Do equilibria exist?

*prices over time (exploration)*

*anticipated price reactions*

*Data-Driven Decision-Making in Enterprise Application*
Further Decision Problems

Revenue Management (Dynamic Programming)
- Inventory Management
- Advertising

Database Configuration (Linear Programming)
- Database Replication
- Data Tiering
- . . .
## Overview

<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 April 25</td>
<td><strong>Linear Programming</strong></td>
</tr>
<tr>
<td>3 April 29</td>
<td>Integer Linear Programming</td>
</tr>
<tr>
<td>4 May 2</td>
<td>Linear/Logistic Regression</td>
</tr>
<tr>
<td>5 May 6</td>
<td>Exercise Implementations</td>
</tr>
<tr>
<td>6 May 16</td>
<td>Dynamic Programming I</td>
</tr>
<tr>
<td>7 May 20</td>
<td>Dynamic Programming II</td>
</tr>
<tr>
<td>8 May 23</td>
<td>Response Strategies / Game Theory</td>
</tr>
<tr>
<td>9 May 27</td>
<td>Project Assignments</td>
</tr>
<tr>
<td>10 June 3</td>
<td>Robust Optimization</td>
</tr>
<tr>
<td>11 June 13</td>
<td>Workshop / Group Meetings</td>
</tr>
<tr>
<td>12 June 20</td>
<td>Presentations (First Results)</td>
</tr>
<tr>
<td>13/14 June 24/27</td>
<td>Workshop / Group Meetings</td>
</tr>
<tr>
<td>15/16 July 1/4</td>
<td>Workshop / Group Meetings</td>
</tr>
<tr>
<td>17 July 11</td>
<td>Presentations (Final Results), Feedback, Documentation (Aug 31)</td>
</tr>
</tbody>
</table>

*Data-Driven Decision-Making in Enterprise Applications - Introduction*