Definitions

Jordan Boyd-Graber
University of Maryland
JANUARY 12, 2018
Roadmap

- What is data science / artificial intelligence / machine learning?
- What is possible with these new tools and techniques?
- What are the challenges?
Data provides the foundation
Machine learning provides algorithms
Artificial intelligence defines problems
Nowhere near true AI
Data are everywhere.
User ratings

**Ikiru** (1952)  
UR  Foreign

**Junebug** (2005)  
R  Independent

**La Cage aux Folles** (1979)  
R  Comedy

**The Life Aquatic with Steve Zissou** (2004)  
R  Comedy

**Lock, Stock and Two Smoking Barrels** (1998)  
R  Action & Adventure

**Lost in Translation** (2003)  
R  Drama

**Love and Death** (1975)  
PG  Comedy

**The Manchurian Candidate** (1962)  
PG-13  Classics

**Memento** (2000)  
R  Thrillers

**Midnight Cowboy** (1969)  
R  Classics
### Purchase histories

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Description</th>
<th>Unit Price</th>
<th>Total Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5/0.51 lb</td>
<td><strong>Cheese</strong>&lt;br&gt;<strong>Cabot Vermont Cheddar</strong>&lt;br&gt;(0.51 lb)</td>
<td>$7.99/lb</td>
<td>$4.07</td>
</tr>
<tr>
<td>1/1</td>
<td><strong>Dairy</strong>&lt;br&gt;<strong>Friendship Lowfat Cottage Cheese</strong>&lt;br&gt;(16oz)</td>
<td>$2.89/ea</td>
<td>$2.89</td>
</tr>
<tr>
<td>1/1</td>
<td><strong>Nature’s Yoke Grade A Jumbo Brown Eggs</strong>&lt;br&gt;(1 dozen)</td>
<td>$1.49/ea</td>
<td>$1.49</td>
</tr>
<tr>
<td>1/1</td>
<td><strong>Santa Barbara Hot Salsa, Fresh</strong>&lt;br&gt;(16oz)</td>
<td>$2.69/ea</td>
<td>$2.69</td>
</tr>
<tr>
<td>1/1</td>
<td><strong>Stonyfield Farm Organic Lowfat Plain Yogurt</strong>&lt;br&gt;(32oz)</td>
<td>$3.59/ea</td>
<td>$3.59</td>
</tr>
<tr>
<td>3/3</td>
<td><strong>Fruit</strong>&lt;br&gt;<strong>Anjou Pears</strong>&lt;br&gt;(Farm Fresh, Med)&lt;br&gt;(1.76 lb)</td>
<td>$2.49/lb</td>
<td>$4.38</td>
</tr>
<tr>
<td>2/2</td>
<td><strong>Cantaloupe</strong>&lt;br&gt;(Farm Fresh, Med)</td>
<td>$2.00/ea</td>
<td>$4.00</td>
</tr>
<tr>
<td>1/1</td>
<td><strong>Grocery</strong>&lt;br&gt;<strong>Fantastic World Foods Organic Whole Wheat Couscous</strong>&lt;br&gt;(12oz)</td>
<td>$1.99/ea</td>
<td>$1.99</td>
</tr>
<tr>
<td>1/1</td>
<td><strong>Garden of Eatin’ Blue Corn Chips</strong>&lt;br&gt;(9oz)</td>
<td>$2.49/ea</td>
<td>$2.49</td>
</tr>
<tr>
<td>1/1</td>
<td><strong>Goya Low Sodium Chickpeas</strong>&lt;br&gt;(15.5oz)</td>
<td>$0.89/ea</td>
<td>$0.89</td>
</tr>
<tr>
<td>2/2</td>
<td><strong>Marcal 2-Ply Paper Towels, 90ct</strong>&lt;br&gt;(1ea)</td>
<td>$1.09/ea</td>
<td>$2.18</td>
</tr>
<tr>
<td>1/1</td>
<td><strong>Muir Glen Organic Tomato Paste</strong>&lt;br&gt;(6oz)</td>
<td>$0.99/ea</td>
<td>$0.99</td>
</tr>
<tr>
<td>1/1</td>
<td><strong>Starkist Solid White Albacore Tuna in Spring Water</strong>&lt;br&gt;(6oz)</td>
<td>$1.89/ea</td>
<td>$1.89</td>
</tr>
</tbody>
</table>
Document collections
Genomics
Social networks

Figure 2: Link communities in US direct flight data detected by Online MMLC. Each segment is less than 500 miles resulting in regional groups. Node sizes on top represent bridgeness, and on the bottom, influence.

MMSB [4] is a Bayesian probabilistic model of network data. MMSB is a mixed membership model. It associates each unit of observation with multiple clusters rather than a single cluster via a membership probability vector. It assumes $K$ communities and directed edges are placed independently between node pairs with probabilities determined only by the pair's community memberships. Given the communities, the generative process defines a multinomial distribution over each node's...
Finance

Dow Jones Composite Average (^DJA) - DJI
4,372.78 ↓ 20.68 (0.47%) 1:48PM EDT

© 2013 Yahoo Inc.
Data are input to algorithms
We can recreate human decisions
We can recreate human decisions
Algorithms can solve tasks
Predictions: Technology changes but society doesn’t
Are we prepared for social changes?
Discussions of AI Need Realism

- You can’t solve a task without good data
- Good data often requires humans

Definitions | UMD
Discussions of AI Need Realism

- You can’t solve a task without good data
- Good data often requires humans
- Unless task is super-straightforward, also need ML expertise
Discussions of AI Need Realism

- You can’t solve a task without good data
- Good data often requires humans
- Unless task is super-straightforward, also need ML expertise
- Effective solutions often require infrastructure
Discussions of AI Need Realism

- You can’t solve a task without good data
- Good data often requires humans
- Unless task is super-straightforward, also need ML expertise
- Effective solutions often require infrastructure
  - Explanations
  - Retraining
  - Socio-technical interfaces
Discussions of AI Need Realism

- You can’t solve a task without good data
- Good data often requires humans
- Unless task is super-straightforward, also need ML expertise
- Effective solutions often require infrastructure
  - Explanations
  - Retraining
  - Socio-technical interfaces
- DS/ML/AI is **hard** (but rewarding!)