Data-Driven Personalization

James Caverlee
Associate Professor
Department of Computer Science and Engineering
Texas A&M University

TAMU BigData Workshop :: 12 February 2016
Our Research:

Web-scale information management
Distributed data-intensive systems
Social computing

Enable efficient and trustworthy information sharing and knowledge discovery over dynamic, heterogeneous and massive-scale networked information systems
Recommended for You

Chrome Browser - Google
FREE
The speed and simplicity of Chrome, now on your Android phone and tablet.

Maps
FREE
Rocky Dunlap, Jeff Korn, Ed Chi, Maayan Roth and Natalie Glance +1’d this

ROM Manager
FREE
Jay Summet, Sashikanth D and Avid Ghamsari +1’d

Google Voice
FREE
Zhijiao Liu, Jeff Korn, Raoul-Sam Daruwala, Ed

Google Play Music
FREE
Jeff Korn, Raoul-Sam Daruwala, Blake Pavel and
## Matrix Completion Perspective

<table>
<thead>
<tr>
<th>User</th>
<th>data mining</th>
<th>anime</th>
<th>fitness</th>
<th>mountain</th>
<th>computer theory</th>
<th>Xbox</th>
</tr>
</thead>
<tbody>
<tr>
<td>User A</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>User B</td>
<td>1</td>
<td>4</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>5</td>
</tr>
<tr>
<td>User C</td>
<td>4</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>5</td>
<td>X</td>
</tr>
</tbody>
</table>

Example from Zhao et al. WWW 2015
Matrix Completion Perspective

<table>
<thead>
<tr>
<th>User</th>
<th>data mining</th>
<th>anime</th>
<th>fitness</th>
<th>mountain</th>
<th>computer theory</th>
<th>Xbox</th>
</tr>
</thead>
<tbody>
<tr>
<td>User A</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>User B</td>
<td>1</td>
<td>4</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>5</td>
</tr>
<tr>
<td>User C</td>
<td>4</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>5</td>
<td>X</td>
</tr>
</tbody>
</table>
Contextual Personalization

• What are informative contextual signals for improving personalization?

• How to exploit and integrate this context into matrix (and tensor) factorization models for improved personalization?

• In particular, how does geo-social context affect personalization?
Research at TAMU: Rise of GeoSocial
Some Perspective on GeoSocial

• Foursquare: over 6 billion check-ins in total. (May 2014)
• Facebook: ~900m photos per day (July 2015)
• Twitter: ~5m geotagged tweets per day (Jan 2013)
• Yelp: ~80 million local reviews (April 2015)
• Smartphones: ~2 billion worldwide by 2016

Our lab: dataset of ~20 billion tweets
Heatmap of List Labelers for Jimmy Fallon
Heatmap of List Labelers of Daniel Vaughn (@BBQsnob)
Tensor-Based Personalized Expert Recommendation using Geo-Topical-Social Context

Figure 1: Overview of the proposed tensor-based personalized expert recommendation framework.
Course Summary

Introduction to the theoretical foundations, algorithms, and methods of deriving valuable insights from data. Includes foundational algorithms; exploratory data analysis; statistical methods and models; and data visualization.

Researchers across disciplines are excited by the prospect of "data-driven science" as a complement to traditional hypothesis-driven "Big Data Research and Development Initiative" spanning NSF, DoD, NIH, DARPA, DoE, and USGS. Companies like Google use analytics to extract information from massive datasets. As a first course in "data science", this course is designed to prepare students for the fields of engineering, statistics, visualization, and experimental design.

Learning outcomes:

- Define and explain the key concepts and models relevant to data science, including data cleaning and integration, data-in-time data visualization.
- Design, implement, and evaluate the core algorithms underlying an end-to-end data science workflow, including the exploration and mining from large datasets.
- Apply "best practices" in data science, including facility with modern tools (e.g., Python, Hadoop, NoSQL databases).