• Introduction
• Data
• Indirect Relationships between Users and Videos
• Evaluation
• Conclusion
Introduction

Love, pretty, work, ...

Sports, Financial, ..., Car

Female

Male
• Mean Girls, Pretty Woman, The Devil Wears Prada
• House of Cards, Mission Impossible, NBA
“Will the **Big Bang Theory** last into the next century?”

“**Sheldon** is so cool, I love him!”

“**Jim Parsons** was nominated for another Emmy Award”
电影: 星球大战7:原力觉醒 2016

导演: J·J·艾布拉姆斯
主演: 哈里森·福特 / 马克·哈米尔

卢卡斯，如是说

明星为什么突然要拍另一部《星球大战》呢？———《星球大战》系列的创始人乔治·卢卡斯在一次采访中回答说：“The ones that got away...”

《星球大战：原力觉醒》彩蛋和花絮汇总

三刷影片后，确认了不少彩蛋，下面尽量按照时间顺序排列。2187 Finn的编号为FN-2187，2187是《新希望》中帝国军克隆 Leia公主的真实编号。这个故事原本就是一个彩蛋，《21-87》是由加拿大先锋导演Arthur...
Introduction

User 1

With the help of

Video name  Actor name  Keyword

User 1

With the help of

Video name  Actor name  Keyword

User 1

With the help of

Video name  Actor name  Keyword

User 1
<table>
<thead>
<tr>
<th>Attribute</th>
<th>Completion Rate</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>95.019%</td>
<td>Male, Female</td>
</tr>
<tr>
<td>Age</td>
<td>18.604%</td>
<td>Teenage (&lt;18), Youngster (18-24), Young (25-34), Mid-age(&gt;34)</td>
</tr>
<tr>
<td>Education BG</td>
<td>17.443%</td>
<td>University, Non-University</td>
</tr>
<tr>
<td>Marital Status</td>
<td>2.203%</td>
<td>Single, Non-Single</td>
</tr>
</tbody>
</table>

**Table 1:** Demographic attributes and corresponding categories
<table>
<thead>
<tr>
<th></th>
<th>Video</th>
<th>Actor</th>
<th>Keyword</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variety show</td>
<td>344</td>
<td>1007</td>
<td>2925</td>
</tr>
<tr>
<td>Movie</td>
<td>306</td>
<td>741</td>
<td>2049</td>
</tr>
<tr>
<td>TV</td>
<td>197</td>
<td>515</td>
<td>1302</td>
</tr>
<tr>
<td>Total</td>
<td>847</td>
<td>1422</td>
<td>4094</td>
</tr>
</tbody>
</table>

**Table 2:** Statics of video relevant information (There is an overlap between the three collections of actors and keywords.)
• Unobvious relationship
  • N*M pairs

• Direct relationship
  • User 2, “Will the Big Bang Theory last into the next century?”

• Indirect relationship
  • User 3 posts, “Sheldon is so cool, I love him!”
Step 1

\[ P(v_n) = \frac{\text{num(\text{users watched the } n_{th} \text{ video})}}{\text{num(\text{users})}} \]

\[ P(w_{ni}|v_n) = \frac{\text{num(\text{users watched the } n_{th} \text{ video and mentioned the } n_{i\text{th}} \text{ keyword})}}{\text{num(\text{users watched the } n_{th} \text{ video})}} \]

\[ P(a_{nj}|v_n) = \frac{\text{num(\text{users watched the } n_{th} \text{ video and mentioned the } n_{j\text{th}} \text{ actor})}}{\text{num(\text{users watched the } n_{th} \text{ video})}} \]

Step 2

\[ P(v_n | W_m, A_k) = \frac{P(W_m, A_k | v_n) * P(v_n)}{P(W_m, A_k)} \]

\[ = \frac{\prod_{w_{ni} \in W_m} P(w_{ni} | v_n) * \prod_{a_{nj} \in A_k} P(a_{nj} | v_n) * P(v_n)}{P(W_m, A_k)} \]
Discover Indirect Relationships

Direct

Direct + Indirect (more denser)

Two Baseline Model

Two Indirect Relationship Based Model

Relationships Tendency

User Video Relationships vs. Iterations
• Matrix Factorization\(^1\), \(K=20\)
• LR\(^2\), SVM\(^2\), GBDT\(^3\)

1 libFFM
2 liblinear
3 XGBoost
• Calculate video demographic tendency
• Calculate user demographic attribute
• Smooth the result
| Feature        | Dis-Baseline | Precision | Recall | F1     | AUC    | Gen-Baseline | Precision | Recall | F1     | AUC    | Gen-Model | Precision | Recall | F1     | AUC    |
|----------------|--------------|-----------|--------|--------|--------|--------------|-----------|--------|--------|--------|--------|-----------|-----------|--------|--------|--------|
| Gender         |              |           |        |        |        |              |           |        |        |        |        |           |           |        |        |        |
|                | Dis-Baseline | 0.720     | 0.714  | 0.717  | 0.730  | Gen-Baseline | 0.701     | 0.687  | 0.694  | 0.707  | Gen-Model | 0.799     | 0.802  | 0.801  | 0.825  |
|                | Dis-Model    | 0.786     | 0.779  | 0.783  | 0.812  | Gen-Baseline | 0.701     | 0.687  | 0.694  | 0.707  | Gen-Model | 0.799     | 0.802  | 0.801  | 0.825  |
|                |              |           |        |        |        |              |           |        |        |        |        |           |           |        |        |        |
| Age            |              |           |        |        |        |              |           |        |        |        |        |           |           |        |        |        |
|                | Dis-Baseline | 0.569     | 0.541  | 0.554  | *      | Gen-Baseline | 0.529     | 0.504  | 0.516  | *      | Gen-Model | 0.663     | 0.645  | 0.654  | *      |
|                | Dis-Model    | 0.642     | 0.653  | **0.648** | **16.8%** | Gen-Baseline | 0.529     | 0.504  | 0.516  | *      | Gen-Model | 0.663     | 0.645  | 0.654  | *      |
|                |              |           |        |        |        |              |           |        |        |        |        |           |           |        |        |        |
| Education BG   |              |           |        |        |        |              |           |        |        |        |        |           |           |        |        |        |
|                | Dis-Baseline | 0.707     | 0.716  | 0.711  | 0.730  | Gen-Baseline | 0.680     | 0.659  | 0.669  | 0.690  | Gen-Model | 0.790     | 0.808  | 0.799  | 0.812  |
|                | Dis-Model    | 0.788     | 0.801  | 0.795  | 0.809  | Gen-Baseline | 0.680     | 0.659  | 0.669  | 0.690  | Gen-Model | 0.790     | 0.808  | 0.799  | 0.812  |
|                |              |           |        |        |        |              |           |        |        |        |        |           |           |        |        |        |
| Marital Status |              |           |        |        |        |              |           |        |        |        |        |           |           |        |        |        |
|                | Dis-Baseline | 0.565     | 0.549  | 0.557  | 0.571  | Gen-Baseline | 0.572     | 0.550  | 0.560  | 0.581  | Gen-Model | 0.682     | 0.691  | 0.687  | 0.696  |

**Table 3:** Prediction accuracy based on users’ video describing words. Classes have been balanced.
Evaluation

(a) AUC of Gender

(b) F1 Score of Age

(c) AUC of Education BG

(d) AUC of Marital Status
Figure 5: Results of Fusion Model evaluation (Macro-F1).
• Our motivation is that user's video related behavior is usually under-utilized on demographic prediction tasks.

• With the help of third-party video sites, we detect the direct and indirect relationships between users and video describing words, and demonstrate this effort can improve the accuracy of users' demographic predictions.

• To our knowledge, this is the first work which explores demographic prediction by fully using users' video describing words.

• This framework has good scalability and can be applied on other concrete features, such as user's book reading behaviors and music listening behaviors.
Thanks!