Entity Recommendations Using Hierarchical Knowledge Bases

Siva Kumar Cheekula, Pavan Kapanipathi, Derek Doran, Prateek Jain, Amit Sheth

Kno.e.sis – Ohio Center of Excellence in Knowledge-enabled Computing
Wright State University, Dayton - OH
Entity Recommendations
Entity Recommendations

Congratulations! Movies we think You will love.
Add movies to your Queue, or Rate ones you've seen for even better suggestions.

Spider-Man 3
300
The Rundown
Bad Boys II
Las Vegas: Season 2
The Last Samurai
Star Wars: Episode III
Robot Chicken: Season 3
(2-Disc Series)
Entity Recommendations
Entity Recommendations
Entity Recommendations

• Content Based (Users’ Ratings)
  – Types
    • Term based
    • Entity based
    • Knowledge based/Graph based
  – Provides the most personalized results

• Collaborative (Other users’ ratings)

• Hybrid (Both content and collaborative)
  – Works best
Knowledge Base Enabled Recommendations

• Structured representation of relationships among concepts

• Allows better understanding of the user interests

• Cold start problem is addressed well
Knowledge Base Enabled Recommendations

Interstellar (film)

Gravity (film)
Knowledge Base Enabled Recommendations

Interstellar (film)

Directed by

Starring

Produced by

Distributed by

Gravity (film)

Relationships
Interstellar (film)
Genre: Science Fiction, Drama, Adventure

Gravity (film)

Directed by
Produced by
Starring
Distributed by

Relationships

Too many properties to address
Hierarchical Relations

- Games
  - Ball Games
    - Football

Is-a

Focus specifically on hierarchical relationships
Hierarchical Relations

• Human memory has been argued to be structured as a hierarchy of concepts (Semantic Network)

• Spreading activation theory has been utilized to simulate search on semantic network

• This theory has not been well explored for user interest modeling and entity recommendations
Wikipedia

- Domain Coverage
- Crowd-sourced knowledge base
  - Regular updates
  - Driven by strict guidelines
Football

From Wikipedia, the free encyclopedia

This article is about the overall concept of games called football. For specific versions of the game, the balls themselves and other uses of the term:

Football refers to a number of sports that involve, to varying degrees, kicking a ball with the foot to score a goal. Unqualified, the word football is understood to refer to whichever form of football is the most popular in the regional context in which the word appears: association football (also known as soccer) in the United Kingdom and most of the non-English speaking world; gridiron football (specifically American football or Canadian football) in the United States and Canada; Australian rules football or rugby league in different areas of Australia; Gaelic football in Ireland; and rugby football (specifically rugby union) in New Zealand.[1][2] These different variations of football are known as football codes.

Categories: Ball games | Football | Sports originating in England | 19th-century introductions
Football

From Wikipedia, the free encyclopedia

This article is about the overall concept of games called football. For specific versions of the game, the balls themselves and other uses of the term.

Football refers to a number of sports that involve, to varying degrees, kicking a ball with the foot to score a goal. Unqualified, the word football is understood to refer to whichever form of football is the most popular in the regional context in which the word appears: association football (also known as soccer) in the United Kingdom and most of the non-English speaking world; gridiron football (specifically American football or Canadian football) in the United States and Canada; Australian rules football or rugby league in different areas of Australia; Gaelic football in Ireland; and rugby football (specifically rugby union) in New Zealand. These different variations of football are known as football codes.
Approach

Wikipedia Category Hierarchy

Spreading Activation

Reverse Spreading

Hierarchical Interest Graphs

Entity Un-rated by user
Hierarchical Interest Graphs

Wikipedia Category Hierarchy

Spreading Activation

Reverse Spreading

Hierarchical Interest Graphs

Entity Un-rated by user

Approach
#semtech talk "How To Make Linked Data More than Data" http://bit.ly/cILgV5
#Semtech #semanticweb #linkeddata federated #Sparql

@google & #semanticweb: Juxtapose 2007 Norvig views/counter views with schema.org: http://bit.ly/google-on-SW
#semtech #schema #semantics

RT @EdGedvilla Great report from The Pew Research Center on "The Fate of the Semantic Web" http://bit.ly/d9ifNw


5 Real world app demoed of semantic sensor web/nw at #ssn2011 #iswc2011 at 3pm

@novas pivak #iswc2009 semantic search w/ ontology driven human assisted site extraction on verts reinvents 2002
http://tinyurl.com/ykg9eqe

A Visualization of the Metadata Universe pdf: http://bit.ly/bOlfsCo via @addthis
#semtech talk "How To Make Linked Data More than Data" http://bit.ly/c1LgV5
#Semtech #semanticweb #linkeddata federated #Sparql

@google & #semanticweb: juxtapose 2007
#semtech #schema #semantics

RT @EdGedvill Great report from The Pew Research Center on "The Fate of the Semantic Web" http://bit.ly/dgifNw


5 Real world app demoed of semantic sensor web/nw at #ssn2011 #iswc2011 at 3pm

@novaspivak #iswc2009 semantic search w/ ontology driven human assisted site extraction on verts reinvents 2002
http://tinyurl.com/ykg9eqe

A Visualization of the Metadata Universe pdf: http://bit.ly/b0IfC0 via @addthis
Hierarchical Interest Graphs

• Wikipedia Category Structure is transformed to a hierarchy (Wikipedia Hierarchy)
  – 0.8 Million Categories with approx 2 Million relationships between them
  – Category “Main Topic Classification” is root of the hierarchy as it subsumes 98% of the categories
  – Each Category is assigned a Hierarchy level based on shortest path from root node

• Spreading Activation Theory on Wikipedia Hierarchy

• Experimented with different Activation Functions
  – Bell, Bell Log, Priority Intersect
Romance Films

Films

Romance Films

Romantic Drama Films

1990s romantic drama films

Titanic
Anna_Karenina
Forever Young

Structured Information

5 4 4
Ratings of Movies

Movies
Romance Films

Titanic
Anna Karenina
Forever Young

1990s romantic drama films

Ratings of Movies

Structured Information
Approach
Ranking Unrated Entities

3
Romantic Drama Films

4
1990s Romantic Drama Films

The Remains of the Day

Jude
Ranking Unrated Entities

Function to rank the entities based on category scores
Reverse Spreading

1984 Films
1980s Action Films
Films Directed by James Cameron
Action Horror Films
Robot Films

The Terminator

1984 Films: 4
1980s Action Films: 3
Films Directed by James Cameron: 2
Action Horror Films: 3
Robot Films: 4
Sum the activation values of each parent of an entity
Summing – Drawbacks

• Ranking is biased towards categories having larger number of entities

• Example:
  – Category “English Language Films” has good interesting score for user
  – Results are topped by the entities connected to “English Language Films” for user
Degree Based Normalization

Out Degree = 3
Spread Weight = 3/3 = 1

- Parent category activation value is normalized by it’s **out degree**
- Bias towards generic categories is addressed
Degree Based Normalization

Out Degree = 3
Spread Weight = 3/3 = 1

- Parent category activation value is normalized by it’s **out degree**
- Bias towards generic categories is addressed
Category Priority

Football

From Wikipedia, the free encyclopedia

This article is about the overall concept of games called football. For specific versions of the game, the balls themselves and other uses of the term:

Football refers to a number of sports that involve, to varying degrees, kicking a ball with the foot to score a goal. Unqualified, the word football is understood to refer to whichever form of football is the most popular in the regional context in which the word appears: association football (also known as soccer) in the United Kingdom and most of the non-English speaking world; gridiron football (specifically American football or Canadian football) in the United States and Canada; Australian rules football or rugby league in different areas of Australia; Gaelic football in Ireland; and rugby football (specifically rugby union) in New Zealand.[1][2] These different variations of football are known as football codes.
Category Priority

- Priority of a category to an entity as edge weight
Reverse Spreading Activation Function

\[ A_i = \Sigma(A_j / (O_j \times P_{ij})) \]

- Activation Value of the Categories of “j”
- Activation Value of the Entity i (Movie)
- Outdegree of Category j
- Priority of Category j to Entity “i”
Evaluation

• Focus on “5” rated movies for test as per Cremonesi et al.
  – New approach for evaluation of recommendation systems

• Dataset - Movielens
  – Movies -- 3158
  – Users -- 1476
  – 0.9 m ratings

• Baseline – Ziegler et al.
  – Hybrid(Content & Collaborative) technique to recommend books using Amazon book taxonomy
  – We compared with the content based part of Ziegler et al
## Recall @Top 1 – Top 20

<table>
<thead>
<tr>
<th>Approach</th>
<th>Summing</th>
<th>Summing with Degree Normalization</th>
<th>Summing with Degree Normalization and Priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decay BellLog</td>
<td>Top 1 -- 0.01</td>
<td>Top 1 -- 0.01</td>
<td>Top 1 -- 0.07</td>
</tr>
<tr>
<td></td>
<td>Top 20 -- 0.15</td>
<td>Top 20 -- 0.18</td>
<td>Top 20 -- 0.21</td>
</tr>
<tr>
<td>Bell Log Intersect</td>
<td>Top 1 -- 0.0</td>
<td>Top 1 -- 0.01</td>
<td>Top 1 -- 0.07</td>
</tr>
<tr>
<td></td>
<td>Top 20 -- 0.09</td>
<td>Top 20 -- 0.20</td>
<td>Top 20 -- 0.24</td>
</tr>
<tr>
<td>Ziegler et al.</td>
<td>Top 1 -- 0.0</td>
<td>Top 1 -- 0.0</td>
<td>Top 20 -- 0.002</td>
</tr>
<tr>
<td></td>
<td>Top 20 -- 0.002</td>
<td>Top 20 -- 0.24</td>
<td></td>
</tr>
</tbody>
</table>
Future Work

• While normalization is necessary, we are looking for better features for normalization

• Priority works, however we find the category-entity priority added by users on Wikipedia needs a quality check

• We are looking into collaborative Hierarchical Interest Graphs
Thank you

Contact
Siva Kumar (siva@knoesis.org)
Pavan Kapanipathi (pavan@knoesis.org) @pavankaps
Amit Sheth (amit@knoesis.org) @amit_p
Cremonesi et al. Top N Evaluation

User Ratings – Test Set

Pick top rated entity ‘T’

Pick 1000 random entities

Generate Rank for 1001 entities

Top N Evaluation

If: Rank(T) <= N
Hits++
Else
Miss++

Recall = Hits/Total