Entity Relatedness Using Explicit Concept Space

Nitish Aggarwal
Insight-Galway, NUIG

Apr 29th, 2014
Entity Relatedness

Entity relatedness is the task of determining the degree of relatedness between two entities.

Brad Pitt  ?  Tom Cruise
Problem Space and Motivation

• Search assistance for exploration in Web Search
  - Some users are short in time and looking for specific answers
  - Other users are willing to explore their interests and extend their knowledge

• Content-based recommender systems
  - User likes some particular items
  - Finding related items which he might like

• Related entities can be obtained by using knowledge bases like DBpedia and Freebase
  - Most of the entities have more than 1000 neighbors and need to be ranked
  - Knowledge bases do not cover all possible related entities for an entity
Problem Space and Motivation

News for brad pitt

LA art fair’s nudes draw close inspection from Brad Pitt on a ...

New York Daily News - 2 days ago
Brad Pitt was surrounded by neked women at Paris Photo Los Angeles. Luckily, for Angelina Jolie, they were just photograhe
the actor was ...

Want to Hang Out With Brad Pitt? Here’s How
ABC News - by Michael Rothman - 3 days ago
Brad Pitt Steps Out Solo for L.A. Art Show—See the Video ...
E! Online - by Marc Malkin - 3 days ago

More news for brad pitt

Brad Pitt - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Brad_Pitt
William Bradley “Brad” Pitt (born December 18, 1963) is an American actor and film producer. He has received a Golden Globe Award, a Screen Actors Guild …
Filmography - List of awards and … - List of most expensive - Douglas Pitt

Brad Pitt - IMDB
www.imdb.com/name/nm0000006/  
Brad Pitt, Actor: Inglourious Basterds. An actor and producer known as much for his versatility as he is for his handsome face. Golden Globe-winning actor Brad …
Awards - Year - Pictures & Photos of Brad Pitt - Rated

Brad Pitt

Actor

William Bradley “Brad” Pitt is an American actor and film producer. Pitt has received five Academy Award nominations and five Golden Globe Award nominations, winning one Golden Globe. Wikipedia

Born: December 18, 1963 (age 50), Shawnee, Oklahoma, United States
Height: 1.80 m
Partner: Angelina Jolie (2005–)
Spouse: Jennifer Aniston (m. 2000–2005)
Children: Shiloh Nouvel Jolie-Pitt, Maddox Chivan Jolie-Pitt, more

Movies

World War 2013 Z
Fight Club 1999
Moneyball 2011
Troy 2004
Killing Them Softly 2012
State of the art

• **Entity relatedness using text relatedness measures**
  - Distributional semantic models such as LSA and ESA
  - Do not distinguish between an entity and a text

• **Entity relatedness using Knowledge bases**
  - Using Wikipedia link structure (Witten and Milne, 2008)
  - Using Wikipedia article’s content (Haffart et. al, 2012)

• **Entity relatedness using search engines features**
  - Using Yahoo query logs and search session history (Blanco et. al, 2013)
  - Using Bing user click logs (Yu et. al, 2014)
Distributional semantics:
- Using unsupervised latent concepts or topics
- Using explicit (manually defined) concepts like Wikipedia concepts
Distributional Semantics for Entity Relatedness (DiSER)

Apple Inc.

1: OS_X (9.31)
3: Apple_Inc. (7.23)
6: ITunes_Store (7.01)
8: MacBook_Pro (6.53)
10: History_of_OS_X (6.44)
11: IPhone (6.32)
12: Goldfrapp_discography (6.24)
15: ITunes (5.87)
17: Apple_TV (5.73)
25: IMac_(Intel-based) (3.56)
30: IPad (3.13)

Steve Jobs

1: Steve_Jobs_(book) (8.21)
3: Apple_Inc. (8.13)
5: History_of_Apple_Inc. (7.43)
7: Macworld/iWorld (7.35)
9: Criticism_of_Apple_Inc. (7.04)
10: IPad (6.98)
14: ITunes (6.27)
17: IPhone (5.43)
22: History_of_the_iPhone (4.13)
23: Macintosh (3.83)
30: OS_X (3.56)

Vector Cosine

Semantic Relatedness
# Distributional Semantics Entity vs Text

## NeXT (Entity)

1: Music_Kit  
2: NeXTSTEP  
3: NeXT_Laser_Printer  
4: NeXT_Computer  
5: Shelf_(computing)  
6: Next  
7: Rich_Page  
8: ISPW  
9: Nancy_R._Heinen  
10: Enterprise_Obъects_Framework  
11: Lotus_Improv  
12: Display_PostScript  
13: NeXTstation  
14: WriteNow  
15: Andrew_C._Stone

## NeXT (Text)

1: Doubly_linked_list  
2: Gare_de_Rennes  
3: Brugge_railway_station  
4: Gare_d'Avignon-Centre  
5: Gare_de_Toulon  
6: Szczecin_Główny_railway_station  
7: Gare_de_Clermont-Ferrand  
8: Leipzig_Central_Station  
9: Brussels-North_railway_station  
10: Gare_de_Strasbourg  
11: Next_friend  
12: Florence_(Amtrak_station)  
13: Reichenau-Tamins_(Rhaetian_Railway_station)  
14: Gare_de_Saint-Pierre-des-Corps  
15: Brussels-South_railway_station
Context-DiSER

Apple Inc.
Apple Inc., formerly Apple Computer, Inc., is an American multinational corporation headquartered in Cupertino, California that designs, develops, and sells consumer electronics, computer software and personal computers. Its best known hardware products are the Mac line of computers, the iPod music player, the iPhone smartphone, and the iPad tablet computer. Its consumer software includes the OS X and iOS operating systems, the iTunes media browser, the Safari web browser, and the iLife and iWork creativity and productivity.

Entity-Linking

Steve Jobs
Steven Paul "Steve" Jobs (/dʒɒbز; February 24, 1955 – October 5, 2011) was an American entrepreneur, marketer, and inventor who was the co-founder (along with Steve Wozniak and Ronald Wayne), chairman, and CEO of Apple Inc. Through Apple, he is widely recognized as a charismatic pioneer of the personal computer revolution and for his influential career in the computer and consumer electronics fields, transforming "one industry after another, from computers and smartphones to music and movies". Jobs also co-founded and served as chief executive of Pixar Animation Studios; he became

Entity-Linking
Context-DiSER

**Apple Inc.**
Apple Inc., formerly Apple Computer, Inc., is an American multinational corporation headquartered in Cupertino, California that designs, develops, and sells consumer electronics, computer software and personal computers. Its best known hardware products are the Mac line of computers, the iPod music player, the iPhone smartphone, and the iPad tablet computer. Its consumer software includes the OS X and iOS operating systems, the iTunes media browser, the Safari web browser, and the iLife and iWork creativity and productivity suites.

**Steve Jobs**
Steven Paul "Steve" Jobs (/dʒəb/) was an American entrepreneur, marketer, and inventor, who was the co-founder (along with Steve Wozniak and Ronald Wayne), chairman, and CEO of Apple Inc. Through Apple, he is widely recognized as a charismatic pioneer of the personal computer revolution and for his influential career in the computer and consumer electronics fields, transforming "one industry after another, from computers and smartphones to music and movies". Jobs also co-founded and served as chief executive of Pixar Animation Studios; he became

**Entity-Context DiSER Vector**

**Entity-Context DiSER Vector**

**Vector Cosine**

**Semantic Relatedness**
Evaluation
Entity Relatedness: Absolute or Relative

• **Absolute relatedness score**
  - Relatedness between “Apple Inc.” and “Steve Jobs”
  - Very low inter-annotator agreement

• **Relative relatedness score**
  - Is “Steve Jobs” is more related with “Apple Inc.” than “Bill Gates”
  - High inter-annotator agreement
 Entity Ranking

1. Steve Jobs
2. Steve Wozniak
3. Jonathan Ive
4. Mac Pro
5. Mike Markkula
6. Infinite Loop (street)
7. Silicon Valley
8. NeXT
9. Safari (web browser)
10. FileMaker Inc.
11. Guy Kawasaki
12. Microsoft
13. Ridley Scott
14. Isaac Newton
15. Alan Turing
16. United States Environmental Protection Agency
17. Greenpeace
18. Ginza
19. Sears
20. Ford Motor Company

KORE Entity Relatedness Gold Standard

21 different entities and every entity has a ranked list of 20 related entities (420 entity pairs in total)
## Entity Ranking: Results

<table>
<thead>
<tr>
<th>Approaches</th>
<th>Spearman Rank Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESA</td>
<td>0.691</td>
</tr>
<tr>
<td>DiSER</td>
<td>0.781</td>
</tr>
<tr>
<td>Context-VSM</td>
<td>0.637</td>
</tr>
<tr>
<td>Context-ESA</td>
<td>0.684</td>
</tr>
<tr>
<td><strong>Context-DiSER</strong></td>
<td><strong>0.769</strong></td>
</tr>
<tr>
<td>KORE (state of art)</td>
<td>0.699</td>
</tr>
</tbody>
</table>
Entity Linking

• Entity linking
  - The task of determining the identity of entities mentioned in text.
  - Steps:
    - Mention identification
    - Candidate retrieval from dictionary
    - Disambiguation

• KORE50 dataset
  - Short context: 14 words per sentence. In contrast, CoNLL-YAGO has ~250 words per document
  - Highly ambiguous mentions: 631 candidates per mentions. In contrast, CoNLL-Yago has only 27.
They performed *Kashmir*, written by *Page* and *Plant*. *Page* played unusual chords on his *Gibson*.

Figure 1: Mention-Entity Graph Example
## Entity Linking: Results

<table>
<thead>
<tr>
<th>Approach</th>
<th>Micro Avg. Precision@1</th>
<th>Macro Avg. Precision@1</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESA</td>
<td>56.8</td>
<td>56.2</td>
</tr>
<tr>
<td>Context-VSM</td>
<td>56.4</td>
<td>55.7</td>
</tr>
<tr>
<td>DiSER</td>
<td>58.3</td>
<td>58.2</td>
</tr>
<tr>
<td>ESA-DiSER</td>
<td>70.2</td>
<td>70.1</td>
</tr>
<tr>
<td>KORE (state-of-the-art)</td>
<td>64.5</td>
<td>62.6</td>
</tr>
</tbody>
</table>
### Entity Relatedness Graph

Please provide Wikipedia link to get related entities.

**Person**

<table>
<thead>
<tr>
<th></th>
<th>Name</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Angelina Jolie (770)</td>
<td>W</td>
</tr>
<tr>
<td>2</td>
<td>George Clooney (929)</td>
<td>W</td>
</tr>
<tr>
<td>3</td>
<td>Jennifer Aniston (559)</td>
<td>W</td>
</tr>
<tr>
<td>4</td>
<td>Tom Cruise (1000)</td>
<td>W</td>
</tr>
<tr>
<td>5</td>
<td>Julia Roberts (640)</td>
<td>W</td>
</tr>
<tr>
<td>6</td>
<td>Johnny Depp (989)</td>
<td>W</td>
</tr>
<tr>
<td>7</td>
<td>Matt Damon (684)</td>
<td>W</td>
</tr>
<tr>
<td>8</td>
<td>Leonardo DiCaprio (742)</td>
<td>W</td>
</tr>
<tr>
<td>9</td>
<td>Bruce Willis (793)</td>
<td>W</td>
</tr>
<tr>
<td>10</td>
<td>Sean Penn (692)</td>
<td>W</td>
</tr>
<tr>
<td>11</td>
<td>Don Cheadle (333)</td>
<td>W</td>
</tr>
<tr>
<td>12</td>
<td>Ben Affleck (648)</td>
<td>W</td>
</tr>
<tr>
<td>13</td>
<td>Tom Hanks (1000)</td>
<td>W</td>
</tr>
<tr>
<td>14</td>
<td>Russell Crowe (647)</td>
<td>W</td>
</tr>
<tr>
<td>15</td>
<td>Cameron Diaz (494)</td>
<td>W</td>
</tr>
<tr>
<td>16</td>
<td>Nicolas Cage (737)</td>
<td>W</td>
</tr>
<tr>
<td>17</td>
<td>Anthony Hopkins (736)</td>
<td>W</td>
</tr>
<tr>
<td>18</td>
<td>Mel Gibson (925)</td>
<td>W</td>
</tr>
<tr>
<td>19</td>
<td>Nicole Kidman (861)</td>
<td>W</td>
</tr>
<tr>
<td>20</td>
<td>Will Smith (1000)</td>
<td>W</td>
</tr>
</tbody>
</table>

**Film**

<table>
<thead>
<tr>
<th></th>
<th>Title</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>World War Z (film) (96)</td>
<td>W</td>
</tr>
<tr>
<td>2</td>
<td>The Curious Case of Benjamin B... (276)</td>
<td>W</td>
</tr>
<tr>
<td>3</td>
<td>Babel (film) (209)</td>
<td>W</td>
</tr>
<tr>
<td>4</td>
<td>Moneyball (film) (159)</td>
<td>W</td>
</tr>
<tr>
<td>5</td>
<td>The Assassination of Jesse Jam... (130)</td>
<td>W</td>
</tr>
<tr>
<td>6</td>
<td>Meet Joe Black (72)</td>
<td>W</td>
</tr>
<tr>
<td>7</td>
<td>Seven (film) (256)</td>
<td>W</td>
</tr>
<tr>
<td>8</td>
<td>The Tree of Life (film) (176)</td>
<td>W</td>
</tr>
<tr>
<td>9</td>
<td>Killing Them Softly (42)</td>
<td>W</td>
</tr>
<tr>
<td>10</td>
<td>Ocean's Twelve (130)</td>
<td>W</td>
</tr>
<tr>
<td>11</td>
<td>Ocean's Thirteen (125)</td>
<td>W</td>
</tr>
<tr>
<td>12</td>
<td>The Favor (22)</td>
<td>W</td>
</tr>
<tr>
<td>13</td>
<td>Fight Club (278)</td>
<td>W</td>
</tr>
<tr>
<td>14</td>
<td>12 Monkeys (150)</td>
<td>W</td>
</tr>
<tr>
<td>15</td>
<td>The Departed (344)</td>
<td>W</td>
</tr>
<tr>
<td>16</td>
<td>Legends of the Fall (107)</td>
<td>W</td>
</tr>
<tr>
<td>17</td>
<td>A River Runs Through It (film) (83)</td>
<td>W</td>
</tr>
<tr>
<td>18</td>
<td>Troy (film) (172)</td>
<td>W</td>
</tr>
<tr>
<td>19</td>
<td>The Dark Knight (film) (683)</td>
<td>W</td>
</tr>
<tr>
<td>20</td>
<td>Megamind (117)</td>
<td>W</td>
</tr>
</tbody>
</table>
Work in Progress

• **Entity Relatedness**
  - Working on temporal distribution
  - Investigating asymmetric relatedness

• **Evaluation**
  - Entity Linking in search query logs
  - DiSER with more LOD based features for content-based RecSys
  - Evaluating DiSER based correlated-topics in text relatedness for cross-lingual scenario
Publications and Participations in 2013-14

Published
“Cross-Lingual Natural Language Querying over the Web of Data” NLDB-2013
“Exploring ESA to Improve Word Relatedness” *SEM-2014

Submitted
“Wikipedia-based Distributional Semantics for Entity Relatedness” ISWC-2014
“Leveraging Structured and Unstructured Context for Medical Concept Resolution” ISWC-2014
“Using Distributional Semantics for Tracing Influence in Romantic-Orientalism Poets”
AHA workshop at Coling-2014

Research Internships
Summer intern at IBM TJ Watson research center, NY
Summer intern at Yahoo! Research Barcelona.

Organizer
1st Insight workshop on Distributional Semantics

External Reviewers
ACL-2014, ISWC-2014, Coling-2014