Search for Complex Objects based on Combination of Attributes and their Affinity

Kosetsu Tsukuda
Kyoto University
tsukuda@dl.kuis.kyoto-u.ac.jp
Background

Increase of task search

20% queries are related to task search [Rosie+ 08]

Task search targeted at complex object

Tourist route in Kyoto
Menu
3 books to read on holiday
I wanna decide four recipes for dinner.

I wanna need two more books to read on holiday.

I wanna delete one spot.
Problem of complex object search (1/2)

It’s not always true that there is an answer in existing complex objects.

Which is an appropriate spot to go with Honkoji and Matsubara-bashi?

There is no tourist routes that include the two spots.

Search

Kyoto tourist route DB

There is no existing combination of objects

There is no appropriate combination of objects
Problem of complex object search (1/2)

If we target **every possible combination** of objects, the number is enormous.

What is tourist route in Kyoto that consists of three spots?

- Need to consider the "appropriateness" of each complex object.
- Include only appropriate complex objects in a search result.
Approach

Focus on two viewpoints that define affinity in the field of psychology. [Winch 58, Byrne 71]

Uniformity (of values)

Complementarity (of personality)

Calculate the affinity b/w objects from the viewpoints of uniformity and complementarity.
Objective

Complex object search based on the affinity of attributes.

Tourist route including three spots in Kyoto

 Unified in the viewpoint of warlord
Sanzenin  Kodaiji  Chisekiin
Hideyoshi Toyotomi  Hideyoshi Toyotomi  Hideyoshi Toyotomi

Complementary in the viewpoint of Buddhism
Tenryuji  Toji  Komyoji
Rinzai sect  Shingon sect  Jodo sect
Definition of symbol

1. \( D \) : Domain
2. \( O_D \) : Complex object
3. \( A_{o_i} \) : Attribute set of \( o_i \in O_D \)
4. \( V_D \) : Viewpoint set
5. \( A_{v_i} \) : Attribute set of \( v_i \in V_D \)

Definition Example

1. \( D = \) Tourist spot in Kyoto
2. \( O_D = \{ \text{Kinkakuji, Toji, ... , Kyoto Tower} \} \)
3. \( A_{\text{Kinkakuji}} = \{ \text{Muromachi era, Rinzai sect, ... , garden} \} \)
4. \( V_D = \{ \text{era, warlord, ... , artist} \} \)
5. \( A_{\text{-era}} = \{ \text{Heian era, Muromachi era, ... , Showa era} \} \)
Definition of uniformity

\( S \subseteq O_D \) meets the uniformity in a viewpoint \( \nu_i \in V_D \).

\[ \nu_i \text{ meets the following two conditions.} \]

1. All objects in \( S \) has **one attribute of viewpoint** \( \nu_i \)
   \[ \forall o_k \in S, |A_{\nu_i} \cap A_{o_k}| = 1 \]

2. The attributes are **same** in all objects.
   \[ |\bigcup_{o_k \in S} A_{\nu_i} \cap A_{o_k}| = 1 \]

**Example**

\( S = \{ \text{Kinkakuji, Ryoanji, Ginkakuji} \} \)

1. All spots have an attribute in a viewpoint “era.”
2. The attribute is “Muromachi” and common.
**Definition of complementarity**

$S \subseteq O_D$ meets complementarity in a viewpoint $\nu_i \in V_D$.

:= $\nu_i$ meets the following two conditions.

1. All objects in $S$ has **one attribute of viewpoint** $\nu_i$

   $$\forall o_k \in S, |A_{\nu_i} \cap A_{o_k}| = 1$$

2. The attributes are **different** in all objects.

   $$|\bigcup_{o_k \in S} A_{\nu_i} \cap A_{o_k}| = |S|$$

**Example**

$S = \{\text{Kinkakuji, Sanzenin, Nanzenji}\}$

1. All spots have an attribute in a viewpoint “religion.”
2. The attributes are different between them like “Rinzai sect,” “Tendai sect,” “Jodo sect.”
Problem definition (search based on the set size)

Input: size $k$ of a complex object

Search based on uniformity

Find $S$ with $v_i \in V_D$ that meets the following conditions.

1. $\forall o_k \in S, |A_{v_i} \cap A_{o_k}| = 1$
2. $|\bigcup_{o_k \in S} A_{v_i} \cap A_{o_k}| = 1$
3. $|S| = k$

Search based on complementarity

Find $S$ with $v_i \in V_D$ that meets the following conditions.

1. $\forall o_k \in S, |A_{v_i} \cap A_{o_k}| = 1$
2. $|\bigcup_{o_k \in S} A_{v_i} \cap A_{o_k}| = |S|$
3. $|S| = k$
Collect attribute etc.

- Information assumed to be already known for problem definition:
  
  1. \(D\) : Domain
  
  2. \(O_D\) : Complex object

  3. \(A_{o_i}\) : Attribute set of \(o_i \in O_D\)

  4. \(V_D\) : Viewpoint set

  5. \(A_{v_i}\) : Attribute set of \(v_i \in V_D\)

- If these pieces of information are identified, the answer sets for each problem are also identified.
Domain name, objects, attributes

Domain name, object

<table>
<thead>
<tr>
<th>Hypernym</th>
<th>Hyponym</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spot in Kyoto</td>
<td>Kinkakuji</td>
</tr>
<tr>
<td>Temple</td>
<td>Kinkakuji</td>
</tr>
<tr>
<td>National univ.</td>
<td>Kyoto univ.</td>
</tr>
</tbody>
</table>

Use open source database\(^1\)

- Domain name
  - hypernym
- Object
  - All hyponyms of the domain

Attributes of an object

Kinkakuji

Extract linked terms in a Wikipedia article

- Few noise terms
- Able to use the open source database for each attribute

\(^1\)http://nlpwww.nict.go.jp/corpus/
Collect viewpoints

All attributes of all objects → Clustering of attributes → Label cluster name

Label = viewpoint
Collect viewpoints

All attributes of all objects  Clustering of attributes  Label cluster name

\[ D = \text{Tourist spot in Kyoto} \]

\[ O_D = \{\text{Kinkakuji, Kiyomizudera, Kyoto Tower, ... , Sanzenin}\} \]

\[ A_{\text{Kinkakuji}} = \{\text{Ukyo – ku, Muromachi era, ... , Rinzai sect}\} \]

\[ A_{\text{Kiyomizudera}} = \{\text{Higashiyama – ku, autumn, ... , the Pillow Book}\} \]

\[ \vdots \]

\[ A_{\text{Sanzenin}} = \{\text{Saityo, prince Moriyoshi, ... , Sakyo – ku}\} \]
Collect viewpoints

All attributes of all objects

Clustering of attributes

Label cluster name
Label = viewpoint

Cluster by furthest neighbor method

- Distance b/w attribute $t_i$ and $t_j$
  \[ d(t_i, t_j) = \frac{1}{\text{#of common hyperny b/w } t_i \text{ and } t_j} \]

- Stopping condition
  \[ \min_{C_i, C_j} D(C_i, C_j) > \theta \]
Collect viewpoints

All attributes of all objects

Clustering of attributes

Label cluster name

Label = viewpoint

Select the most informative hypernym as the label.

Common hypernyms of all Attributes In the left cluster.

- Male
- Eastern artist

# of attribute whose hypernym is “Eastern artist” < # of attribute whose hypernym is “Male”

∴ “Eastern artist” is more informative
<table>
<thead>
<tr>
<th>Domain</th>
<th>Tourist spot in Kyoto city</th>
</tr>
</thead>
<tbody>
<tr>
<td># of objects</td>
<td>168</td>
</tr>
<tr>
<td>Average # of attributes</td>
<td>73</td>
</tr>
<tr>
<td># of clusters (viewpoints)</td>
<td>436</td>
</tr>
<tr>
<td>Average cluster size</td>
<td>4.1</td>
</tr>
</tbody>
</table>
## Result (search based on set size)

### Unified search result (set size: 3)

<table>
<thead>
<tr>
<th>Viewpoint</th>
<th>Object (attribute)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Style of main shrine</td>
<td>Sanjusangendo (hipped roof), Daikakuji (hipped roof), Daitokuji (hipped roof)</td>
</tr>
<tr>
<td>Buddhism</td>
<td>Higashihonganji (Jodo sect), Chionin (Jodo sect), Nishihonganji (Jodo sect)</td>
</tr>
<tr>
<td>City</td>
<td>Myoshinji (Kyoto city), Syokokuji (Kyoto city), Chionin (Kyoto city)</td>
</tr>
</tbody>
</table>

### Complementary search result (set size: 3)

<table>
<thead>
<tr>
<th>Viewpoint</th>
<th>Object (attribute)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eastern artist</td>
<td>Mibudera (Tohaku hasegawa), Nanzenji (Tanyu Kano), Daisenin (Motonobu kano)</td>
</tr>
<tr>
<td>Warlord</td>
<td>Kiyomizudera (Hideyoshi Toyotomi), Tojiin (Masanori Fukushima), Jojakkoji (hideaki Kobayakawa)</td>
</tr>
</tbody>
</table>
Result (search for additional objects)

**Unified** search result
- Original complex object: \{Tenryuji, Toji, Shogoin\}
- Additional quantity: 2

<table>
<thead>
<tr>
<th>Viewpoint</th>
<th>Object (attribute)</th>
</tr>
</thead>
<tbody>
<tr>
<td>People in Muromachi era</td>
<td>Tenryuji (Emperor Go-Daigo), Toji (Emperor Go-Daigo), Syogoin (Emperor Go-Daigo), <strong>Bukkoji (Emperor Go-Daigo)</strong>, <strong>Myokenji (Emperor Go-Daigo)</strong></td>
</tr>
</tbody>
</table>

**Complementary** search result
- Original complex object: \{Kinkakuji, Ninnaji\}
- Additional quantity: 1

<table>
<thead>
<tr>
<th>Viewpoint</th>
<th>Object (attribute)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japan movie</td>
<td>Kinkakuji (Goban-cho Yugiri-ro), Ninnaji (I Give My First Love to You), <strong>Fushimi Inari Taisha (Owl’s castle)</strong></td>
</tr>
</tbody>
</table>
Result (search for deducted objects)

Unified search result
- Original complex object: \{Ninnaji, Tofukuji, Ginkakuji\}
- Deducted quantity: 1

<table>
<thead>
<tr>
<th>Viewpoint</th>
<th>Object (attribute)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anime</td>
<td>Ninnaji (Detective Conan: Crossroad in the Ancient Capital), Tofukuji (Detective Conan: Crossroad in the Ancient Capital)</td>
</tr>
</tbody>
</table>

Complementary search result
- Original complex object: \{Sanzenin, Fushimi Inari Taisha, Nanzenji, Daigoji, Rokuonji\}
- Deducted quantity: 1

<table>
<thead>
<tr>
<th>Viewpoint</th>
<th>Object (attribute)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buddhist sect</td>
<td>Sanzenin (Tendai sect), Fushimi Inari Taisha (Shingon sect), Nanzenji (Jodo sect), Rokuonji (Rinzai sect)</td>
</tr>
</tbody>
</table>
• There are **too many combinations** to show all of them.
  – E.g. combination of three tourist spots:
    • Total complex objects: 776,216 combinations.
    • Unified complex objects: 197,649 combinations.
    • Complementary complex objects: 24,608 combinations.

• Need to **rank** complex objects.
  – Place a complex object whose viewpoint is highly relevant to the domain in high ranking.
    • E.g. “Eastern artist” is more relevant to “tourist route in Kyoto” than “anime.”
Conclusion

Objective

Complex object search based on the affinity of attributes.

Focus on **uniformity** and **complementarity** b/w objects.

Future work

- Use other resources for extracting attributes.
- Apply other domains such as novel.
- Evaluation.