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

Target search intent

Search based on the set size.

> I wanna decide four recipes for dinner.

Search additional objects.

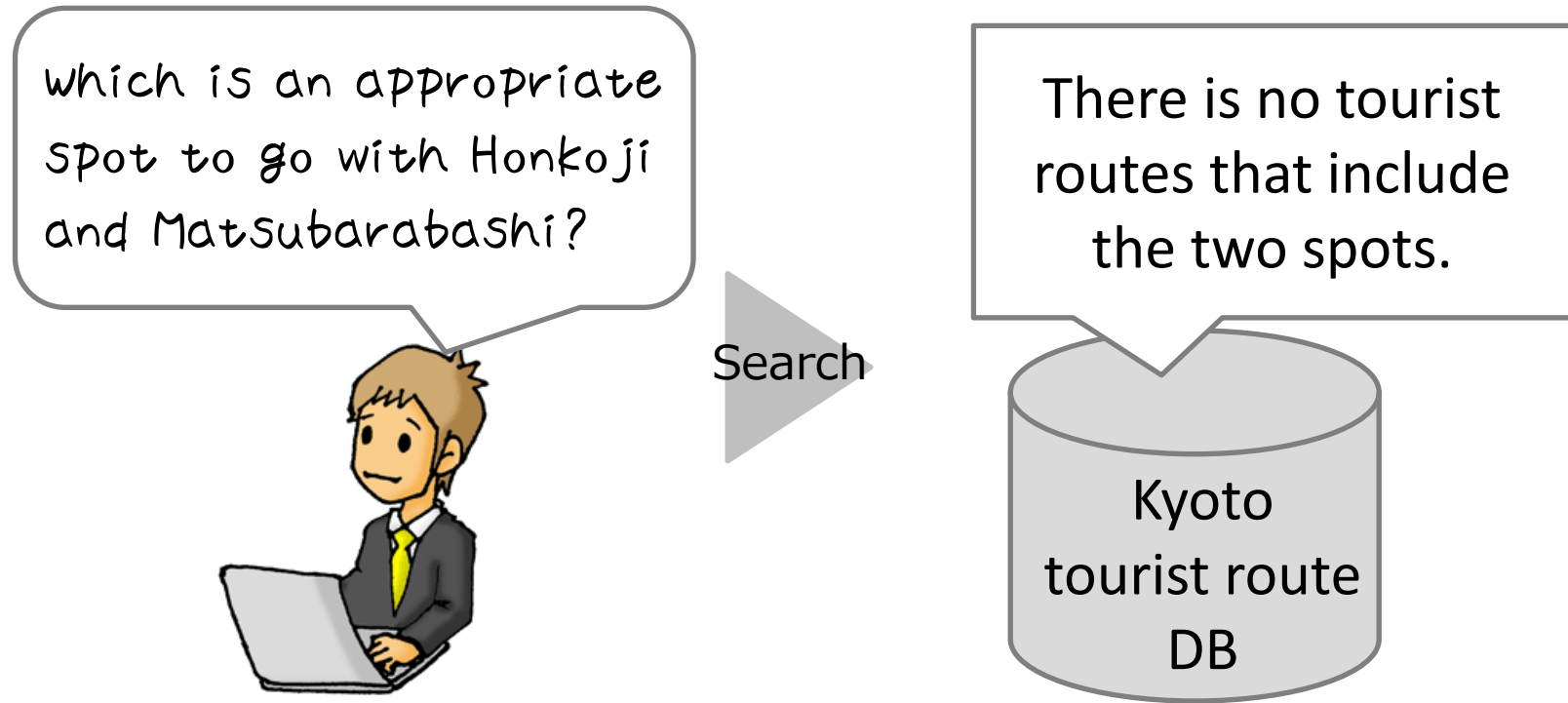> I wanna need two more books to read on holiday.

Add

Search deducted objects.

> I wanna delete one spot.

Delete

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

Which is an appropriate spot to go with Honkoji and Matsubarabashi?

Search

There is no tourist routes that include the two spots.

Kyoto tourist route DB

There is no existing combination of objects
⊬
There is no appropriate combination of objects

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

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

Possible combination

1000000 Hit!

- 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]

Results-oriented   Results-oriented

Shy   Outgoing

Uniformity
(of values)

Complementarity
(of personality)

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

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

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 $v_i \in V_D$

Example

1. D = Tourist spot in Kyoto

2. O_D = {Kinkakuji, Toji, … , Kyoto Tower}

3. $A_{\text{Kinkakuji}}$ = {Muromachi era, Rinzai sect, … , garden}

4. V_D = {era, warlord,…, artist}

5. A_{era} = {Heian era, Muromachi era,…,Showa ear}

Definition of uniformity

$S \subseteq O_D$ meets the uniformity in a viewpoint $v_i \in V_D$.

$:= v_i$ meets the following two conditions.

1. All objects in S has **one attribute of viewpoint v_i**

$$\forall o_k \in S, \left| A_{v_i} \cap A_{o_k} \right| = 1$$

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

$$\left| \bigcup_{o_k \in S} A_{v_i} \cap A_{o_k} \right| = 1$$

Example

$S = \{\text{Kinkakuji}, \text{Ryoanji}, \text{Ginkakuji}\}$

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 $v_i \in V_D$.

$:= v_i$ meets the following two conditions.

1. All objects in S has **one attribute of viewpoint** v_i

$$\forall o_k \in S, \left| A_{v_i} \cap A_{o_k} \right| = 1$$

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

$$\left| \bigcup_{o_k \in S} A_{v_i} \cap A_{o_k} \right| = |S|$$

Example

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

1. All spots have an attribute in a viewpoint "religion."
2. The attributes are different between them like "Rinzai sect," "Tendai sect," "Jodo sect."

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

Search based on complementarity

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

1. $\forall o_k \in S, \left| A_{v_i} \cap A_{o_k} \right| = 1$
2. $\left| \cup_{o_k \in S} A_{v_i} \cap A_{o_k} \right| = |S|$
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

Hypernym	Hyponym
Spot in Kyoto	Kinkakuji
Temple	Kinkakuji
National univ.	Kyoto univ.
⋮	⋮

Use open source database[1]
- Domain name
  - hypernym
- Object
  - All hyponyms of the domain

[1]http://nlpwww.nict.go.jp/corpus/

Attributes of an object

Kinkakuji
────────
────────────
─ Ukyoku ──────
────────────
──── Muromachi ·
─ Ashikaga ─────
────────────

Extract linked terms in a Wikipedia article
- Few noise terms
- Able to use the open source database for each attribute

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**
Label＝viewpoint

D = Tourist spot in Kyoto

$O_D = \{\text{Kinkakuji, Kiyomizudera, Kyoto Tower, } \ldots \text{, Sanzenin}\}$

$A_{\text{Kinkakuji}} = \{\text{Ukyo} - \text{ku, Muromachi era, } \ldots \text{, Rinzai sect}\}$

$A_{\text{Kiyomizudera}} = \{\text{Higashiyama} - \text{ku, autumn, } \ldots \text{, the Pillow Book }\}$

$\vdots$

$A_{\text{Sanzenin}} = \{\text{Saityo, prince Moriyoshi, } \ldots \text{, Sakyo} - \text{ku}\}$

Collect viewpoints

**All attributes of
all objects** ▶ **Clustering of
attributes** ▶ **Label cluster name**
Label＝viewpoint

Shingon

Kegon

Rinzai

Ritu

Jodo

Tendai

Sodatsu Tawaraya

Tanyu kano

Motonobu kano

Ukyo

Sakyo

Tohaku Hasegawa

Higashiyama
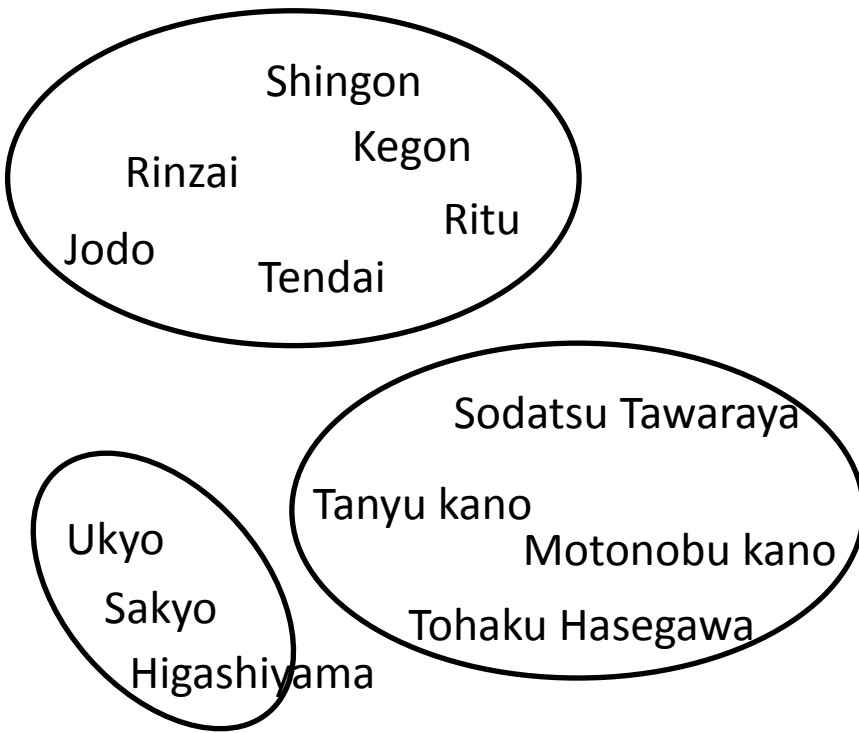
Cluster by furthest neighbor method

– 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
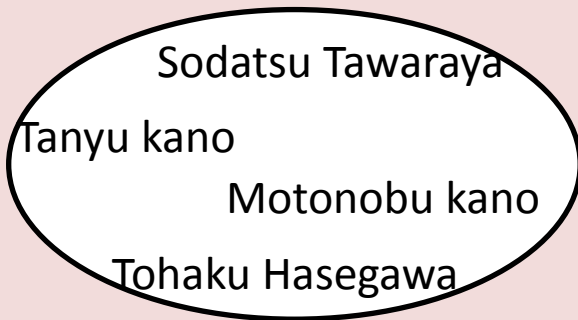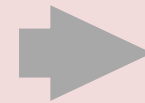
All attributes of all objects ▶ **Clustering of attributes** ▶ **Label cluster name**
Label ＝ viewpoint

Select the most informative hypernym as the label.

Sodatsu Tawaraya
Tanyu kano
Motonobu kano
Tohaku Hasegawa

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

Experiment

Domain	Tourist spot in Kyoto city
# of objects	168
Average # of attributes	73
# of clusters (viewpoints)	436
Average cluster size	4.1

Result (search based on set size)

Unified search result (set size: 3)

Viewpoint	Object (attribute)
Style of main shrine	Sanjusangendo (hipped roof), Daikakuji (hipped roof), Daitokuji (hipped roof)
Buddhism	Higashihonganji (Jodo sect), Chionin (Jodo sect), Nishihonganji (Jodo sect)
City	Myoshinji (Kyoto city), Syokokuji (Kyoto city), Chionin (Kyoto city)

Future task

Complementary search result (set size: 3)

Viewpoint	Object (attribute)
Eastern artist	Mibudera (Tohaku hasegawa), Nanzenji (Tanyu Kano), Daisenin (Motonobu kano)
Warlord	Kiyomizudera (Hideyoshi Toyotomi), Tojiin (Masanori Fukushima), Jojakkoji (hideaki Kobayakawa)

Future Task

An attribute that is common in many objects also appears.

Result (search for additional objects)

Unified search result

- Original complex object: {Tenryuji, Toji, Shogoin}
- Additional quantity: 2

Viewpoint	Object (attribute)
People in Muromachi era	Tenryuji (Emperor Go-Daigo), Toji (Emperor Go-Daigo), Syogoin (Emperor Go-Daigo), **Bukkoji (Emperor Go-Daigo), Myokenji (Emperor Go-Daigo)**

Complementary search result

- Original complex object: {Kinkakuji, Ninnaji}
- Additional quantity: 1

Viewpoint	Object (attribute)
Japan movie	Kinkakuji (Goban-cho Yugiri-ro), Ninnaji (I Give My First Love to You), **Fushimi Inari Taisha (Owl's castle)**

Diversity of viewpoints reduced as the set size becomes bigger.

Result (search for deducted objects)

Unified search result

- Original complex object: {Ninnaji, Tofukuji, Ginkakuji}
- Deducted quantity: 1

Viewpoint	Object (attribute)
Anime	Ninnaji (Detective Conan: Crossroad in the Ancient Capital), Tofukuji (Detective Conan: Crossroad in the Ancient Capital)

Complementary search result

- Original complex object: $\left\{ \begin{array}{c} \text{Sanzenin, Fushimi Inari Taisha,} \\ \text{Nanzenji, Daigoji, Rokuonji} \end{array} \right\}$
- Deducted quantity: 1

Viewpoint	Object (attribute)
Buddhist sect	Sanzenin (Tendai sect), Fushimi Inari Taisha (Shingon sect), Nanzenji (Jodo sect), Rokuonji (Rinzai sect)

Future work

- 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.