NTCIR

The New INTENT

From INTENT2 to ~~INTENT3~~

Yiqun Liu, Ruihua Song, Min Zhang, Zhicheng Dou,
Takehiro Yamamoto, Makoto P. Kato, Hiroaki Ohshima

June, 2013@NTCIR-10, Tokyo

Lessons from Past INTENT Tasks

- About result diversification:
  - INTENT@NTCIR9: Top performers may be either relevance-oriented or diversity-oriented
  - INTENT2@NTCIR10: Diversified ranking results significantly outperformed baseline non-diversified run.
- About the task settings:
  - SM task attracts much more participants than DR
  - Subtopic annotation (clustering and importance voting) is rather difficult even for organizers and professional assessors.
  - Whether ordinary users are more satisfied with diversified search results remains unknown.

NTCIR

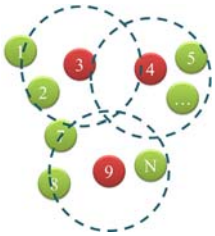
More User Behavior Data

- THUIR@INTENT/INTENT2: More user behavior data led to better performance.
- More raw click-through data: SogouQ has doubled its size to include click-through data collected from Sogou.com in 2011
  - 1.85GB => 3.85GB, over 40M user clicks
- More subtopic candidates generated from more recent user behavior data.
  - Search engine data provided by two major Chinese search engines will be adopted
- Find out whether more logs help improve SM/DR performance

NTCIR

Intent Annotation Using Logs

- SM results in pools will be clustered with click-through/pseudo RF data at first to generate preliminary candidate intent groups.
- Query frequency information will be taken into consideration during subtopic importance voting process.
- Data source: recently collected data from Sogou for Chinese SM, Bing for English/Japanese SM
- More credible SM qrels with less annotation efforts
- Perhaps the reusability of results could be increased



NTCIR

Crowd-Sourcing based Evaluation

- A search-engine-like annotation interface to collect feedback information from a relatively large number of unprofessional users (e.g. 50+ undergraduate students)
- Data collected from the interface: query, click, examination

$$P(C_i = 1) = P(E_i = 1)P(R_i = 1)$$

- High correlation with examination
  - Preliminary results on 10 users.
  - KAPPA: 0.65, Accuracy: 0.83.
- Find out whether D# measures accords with user satisfaction

INTENT with/for Knowledge Graph

- Fuji in Wikipedia:
  - Fuji the actress
  - Fuji the Mountain
  - Fuji the apple

- Ambiguous queries: NTCIR INTENT #0205: 功夫(kung fu)，#0206: 生日快樂(happy birthday)
- INTENT task with existing knowledge graphs such as wikipedia, freebase (e.g. THCIB and THUIS in INTENT2)
- INTENT task to help improve existing knowledge graphs

Summary of Proposal

	INTENT2	*The NEW INTENT*
Number of Topics	• Chinese: 100 • Japanese: 100 • English: 50	• Chinese: 100 • Japanese: 100 • English: 100
DR task setting	• Chinese: SogouT (Ver.2008) • Japanese: ClueWeb JA	• Chinese: SogouT (Ver.2012) • English: ClueWeb12-B13
Support from log analysis for annotation	No	Support from log analysis for SM annotation
Crowd sourcing	No	Crowd sourcing for Chinese DR
Subtopic candidate	Query suggestions from Bing, Google, Sogou and Baidu	• Query suggestions from Bing, Google, Sogou and Baidu • Candidates generated by MSRA from search engine result page • Candidates generated by THU from Sogou log data
User behavior data	SogouQ (data collected in 2008): 2GB approximately	SogouQ (data collected in 2008 and 2011): 4GB approximately
DR Baseline	• Chinese DR baseline • Japanese DR baseline	• ClueWeb12-B13 retrieval service is provided by CMU • SogouT retrieval service is provided by THU/SOGOU

NTCIR

Task Mining Task (TASKMINE) Task Proposal

- As the NEW INTENT's Subtask -

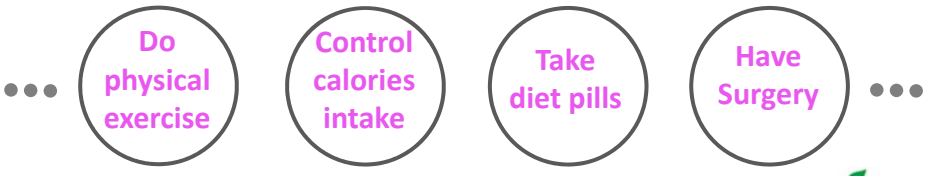
Takehiro Yamamoto
Makoto P. Kato
Hiroaki Ohshima
(Kyoto University)

June, 2013@NTCIR-10, Tokyo

Overview

Given a query, find the set of tasks to achieve it

Lose Weight

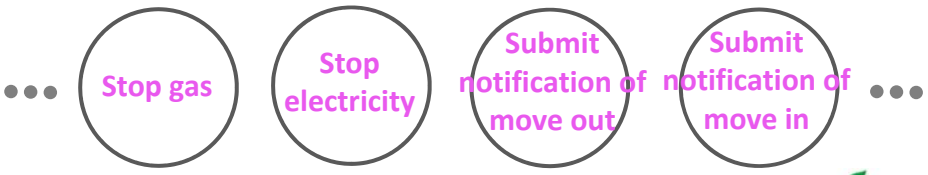


NTCIR

Overview

Given a query, find the set of tasks to achieve it

Move to a new house



NTCIR

Why Important?

Demand for Supporting Complex Search Task

- Domains such as Health, Education, Travel are tends to be complex tasks, and should be supported [Donate *et al.* WWW2010]
- Subgoal mining from Sponsored search data [Yamamoto *et al.* CIKM2012]
- Grouping relevant tasks by ODP Categories [Hassan *et al.* CIKM2012]

- Yamamoto, T., Sakai, T., Iwata, M., Chen, Y., Wen, J.-R. and Tanaka, K.: The Wisdom of Advertisers: Mining Subgoals via Query Clustering, ACM CIKM 2012, pp. 505-514, 2012.
- Hassan, A. and White, R. W.: Task Tours: Helping Users Tackle Complex Search Tasks, ACM CIKM 2012, pp. 1185-1189, 2012.

NTCIR

Our Task Model

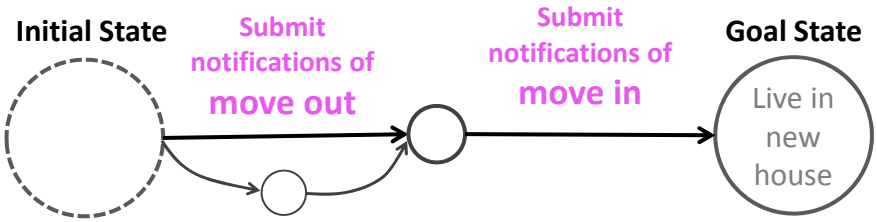
Task is a set of **actions** that approach to the **goal state**

NTCIR

Our Task Model

Task is a set of **actions** that approach to the **goal state**

Move to a new house



NTCIR

Possible Research Issues

- IR People
  - iUnit (information unit) based retrieval (1CLICK-2)
  - Subtopic mining techniques (INTENT-2)
- NLP People
  - Procedural information extraction
  - Entailment relation extraction (RITE), especially verb entailment
- QA People
- And more

NTCIR

Task Plans

- Task Settings
  - Manually prepare 100 queries (tasks)
  - Participants' output format:
    Ranked list of tasks with their temporal orders
- Language
  - Japanese
- Evaluation Method
  - Manually Create gold-standards
    - Annotate task relationships
      (temporal, part-of, is-a, hierarchical, etc.)
  - Apply new evaluation metrics (hierarchical, and more)

NTCIR

Summary

TASKMINE as INTENT-3 Subtask

Given a query, find the set of **tasks** to achieve it

Research Challenges:

- How to **Retrieve** tasks from the Web?
- How to **Model** relationships among tasks?

NTCIR

Thank you

NTCIR