Anticipating Information Needs Based on Check-in Activity

Author: Jan R. Benetka , Krisztian Balog , Salvatore Giorgi, Kjetil Nørvåg
Source: WSDM’17
Speaker: 游雅雯
Date: 2017/6/9

Outline

- Introduction
- (RQ1)Information Needs Related to Activities
- (RQ2)Analysis of Temporal Dynamics of Information Needs
- (RQ3)Anticipating Information Needs
- Conclusion

Introduction

Zero-query Search

- using the person's check-in activity as context
- anticipate information needs
- respond with a set of information cards

Introduction

Category is a very strong indicator of human activity

An ***activity***, in the scope of this paper, is defined as a category of a point-of-interest (POI) that the user visited.

Introduction

<u>Research questions</u>

RQ1) What are common information needs and their relevance in the context of different activities?

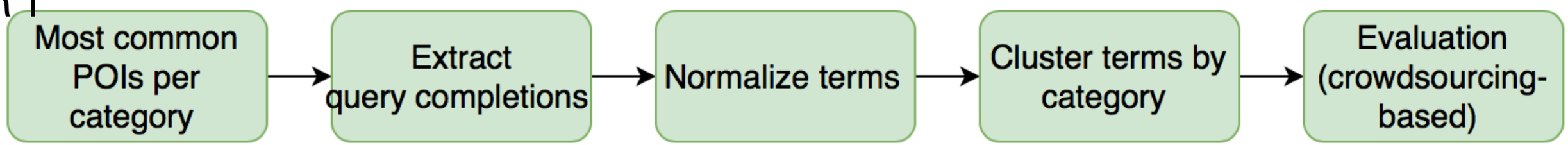
RQ2) Are information needs static or dynamic in time?

RQ3) How to rank future information needs given the last activity of the user?

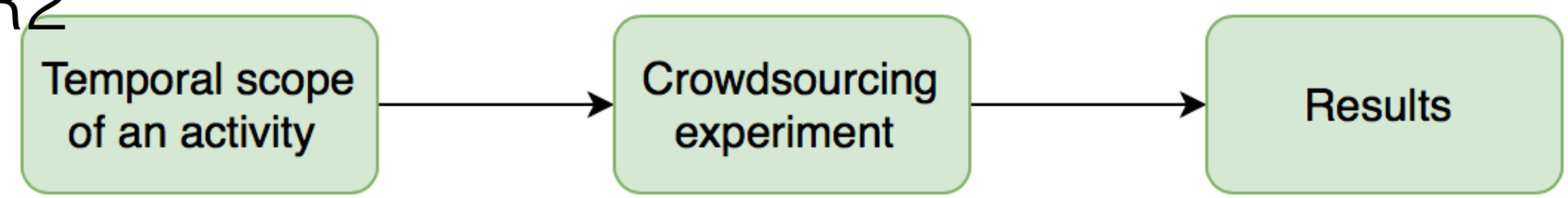
Introduction

Flow Chart

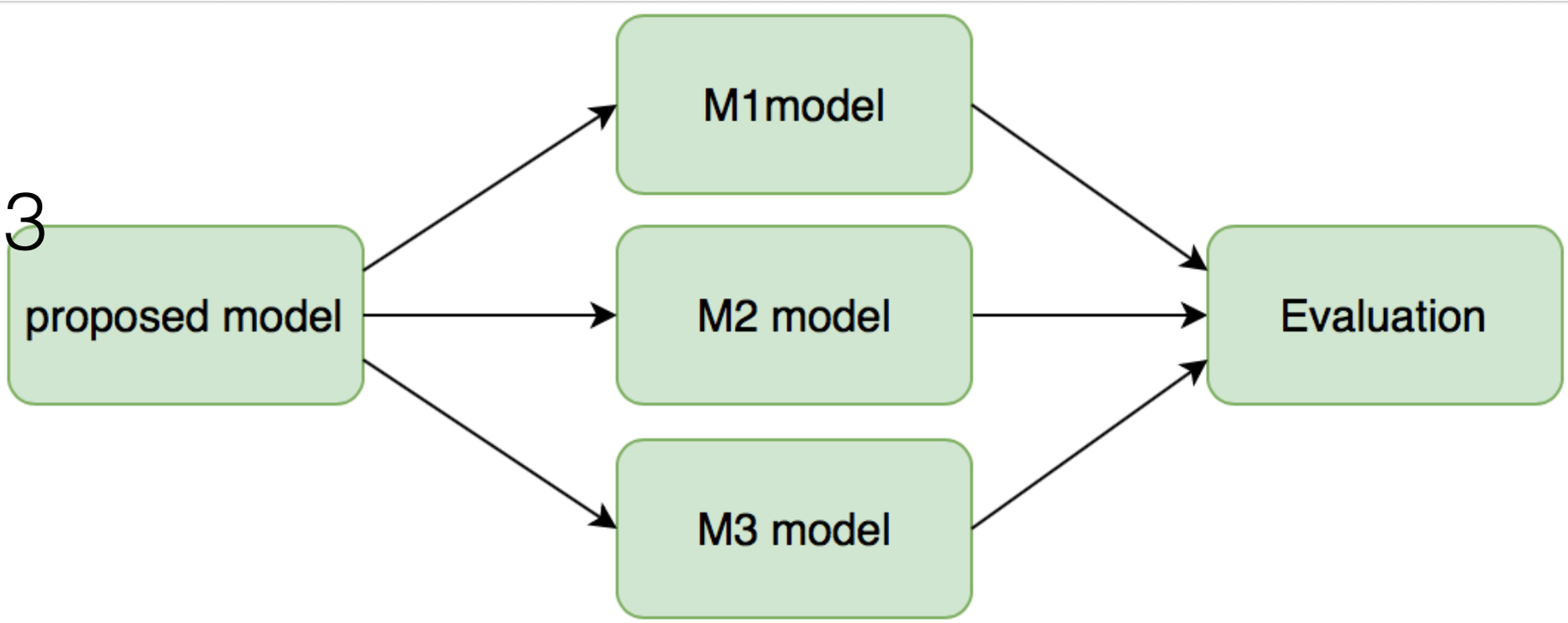
R1



R2



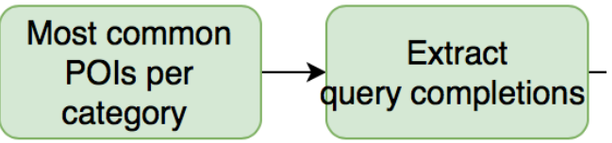
R3



Outline

- Introduction
- <u>(RQ1)Information Needs Related to Activities</u>
- (RQ2)Analysis of Temporal Dynamics of Information Needs
- (RQ3)Anticipating Information Needs
- Conclusion

Information Needs Related to Activities(RQ1)



Check-in data

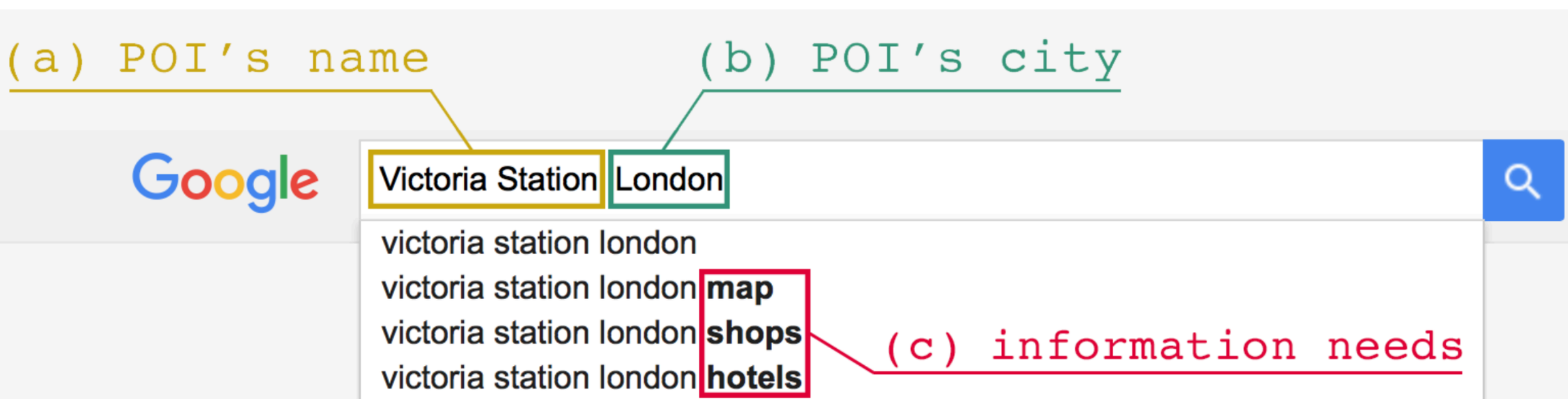
- Foursquare Category Hierarchy
- Foursquare API

Collecting Query Suggestions

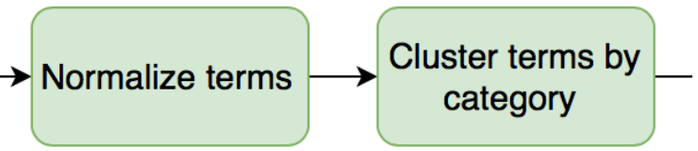
- Google Query Suggestion API

Category: **Metro Station**

POIs:
- **Victoria**
- Waterloo
- King's Cross St. Pancras
- Oxford Circus
- ...



Information Needs Related to Activities(RQ1)



Normalization

- graph-based approach
- DPClus graph clustering algorithm

Determining relevance

$$P(i|a) = \frac{n(i,a)}{\sum_{i' \in I} n(i',a)}$$

Table 2: Information need labels and their synonym terms.

Information need	Synonyms
jobs	employment, job, careers, career, . . .
map	localization map, map, travel maps, . . .
prices	price list, price, prices, costs, taxi rate, . .
operation hours	opening time, office hours, times, . . .

Information Needs Related to Activities(RQ1)

Evaluation (crowdsourcing-based)

Evaluation

- compare the extracted information needs against crowdsourced human judgments

Category	#Needs	R@10	R@20	R@All
College & University	27	0.22	0.37	0.74
Food	15	0.53	0.53	0.73
Residence	36	0.22	0.25	0.28
Travel & Transport	25	0.24	0.36	0.48
Outdoors & Recreation	19	0.26	0.53	0.89
Arts & Entertainment	22	0.23	0.27	0.68
Shop & Service	22	0.23	0.36	0.77
Nightlife Spot	18	0.33	0.50	0.78
Professional & Other Places	31	0.26	0.39	0.65
Average	23.9	0.28	0.40	0.67

Information Needs Related to Activities(RQ1)

Evaluation (crowdsourcing-based)

Evaluation

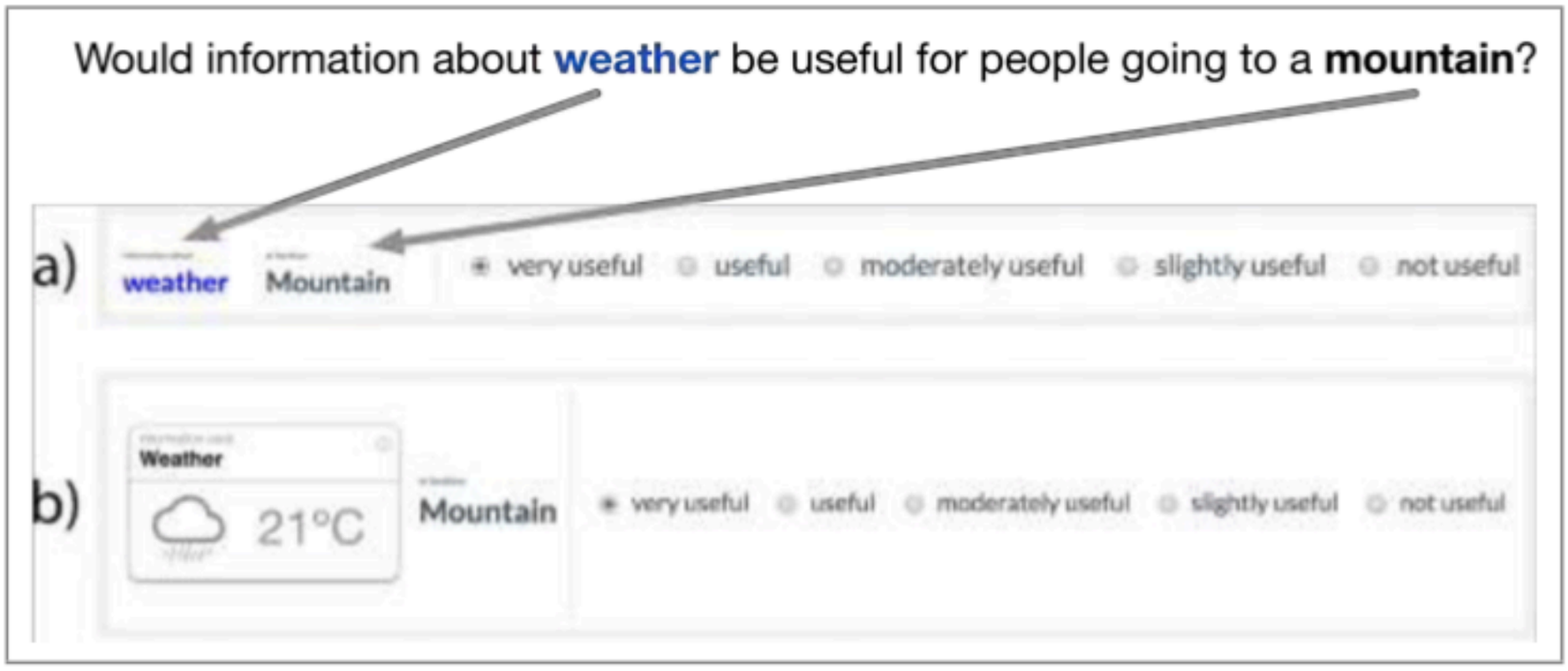


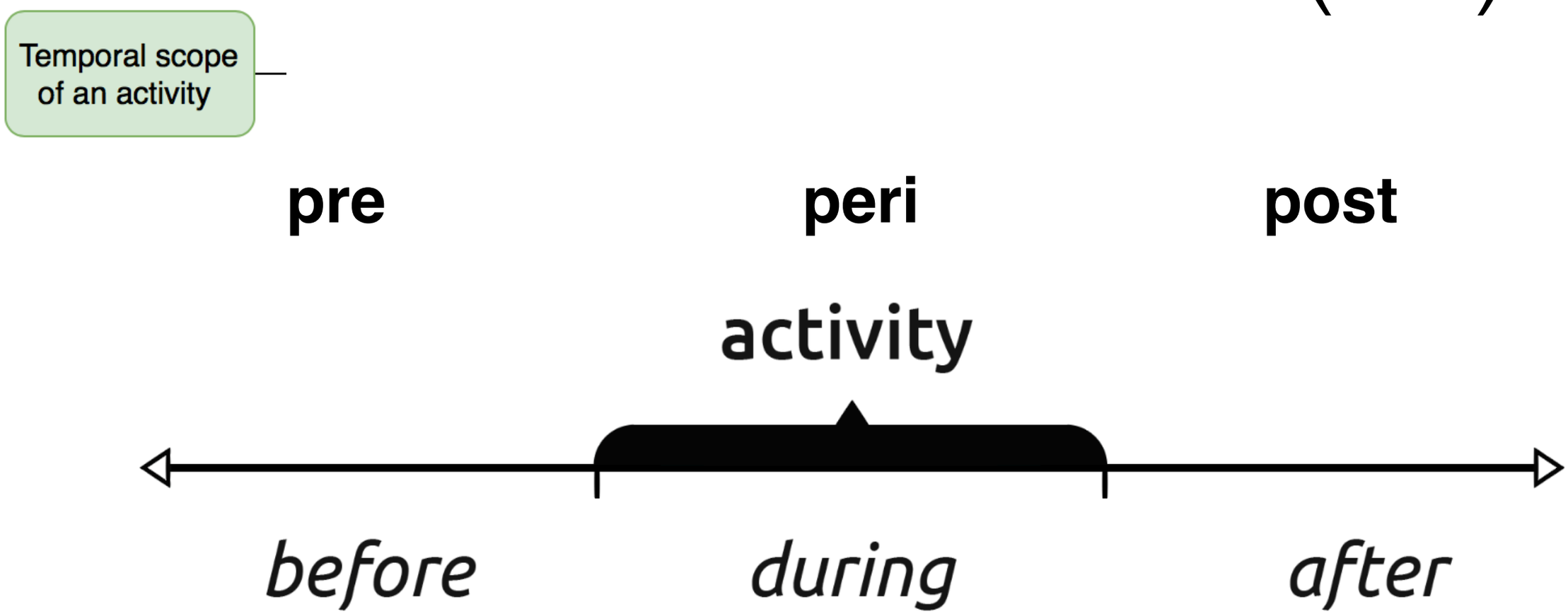
Table 4: Evaluation of the ranking of information needs with respect to their relevance for a given activity.

Ground truth	NDCG@3	NDCG@5	NDCG@10
Text-based	0.491	0.550	0.627
Card-based	0.519	0.535	0.603

Outline

- Introduction
- (RQ1)Information Needs Related to Activities
- <u>(RQ2)Analysis of Temporal Dynamics of Information Needs</u>
- (RQ3)Anticipating Information Needs
- Conclusion

Analysis of Temporal Dynamics of Information Needs(R2)



Temporal scope

$$P(t|i,a) = \frac{n(t,i,a)}{\sum_{t \in \{pre,peri,post\}} n(t,i,a)}$$

Analysis of Temporal Dynamics of Information Needs(R2)

Crowdsourcing experiment

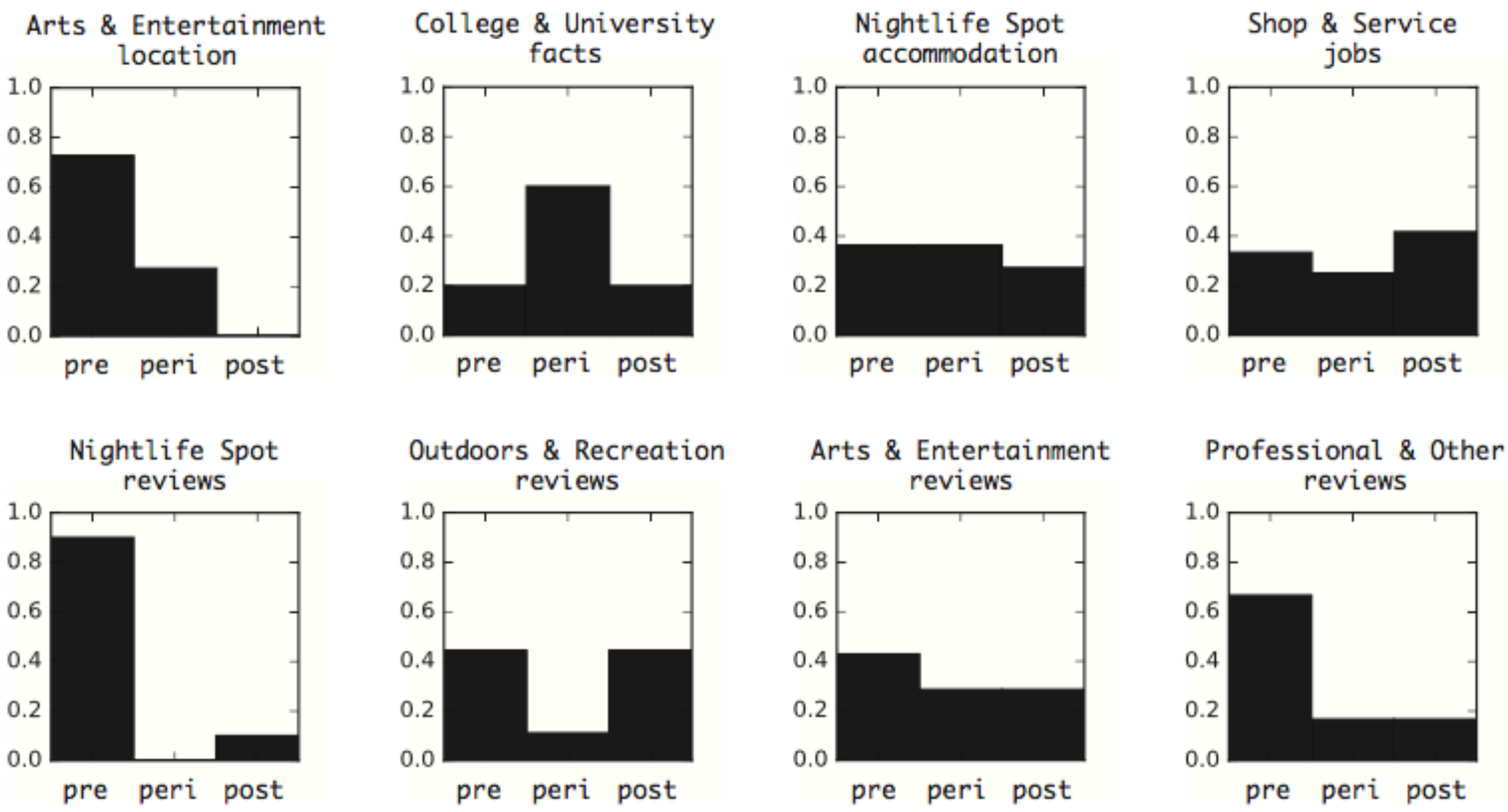
Category: Restaurant

location	✓ *before*	○ *during*	○ *after*
menu	✓ *before*	✓ *during*	○ *after*
review	✓ *before*	○ *during*	✓ *after*

Analysis of Temporal Dynamics of Information Needs(R2)

Evaluation

Evaluation



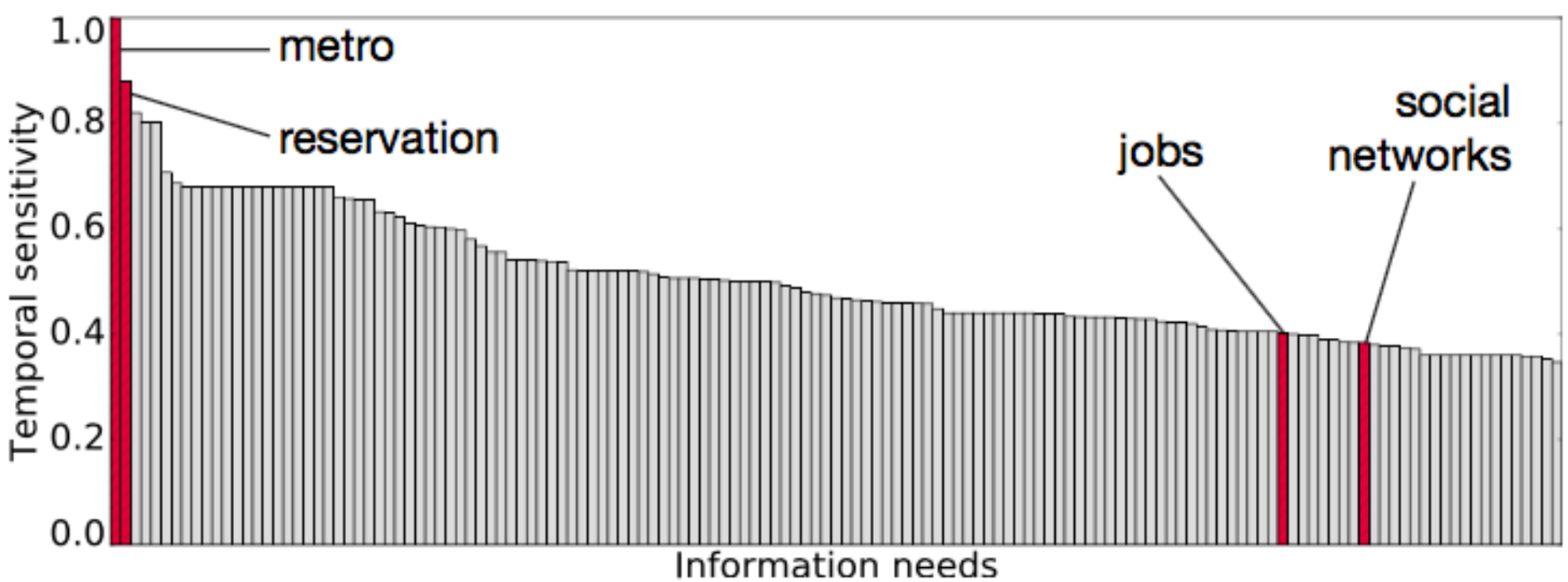
Analysis of Temporal Dynamics of Information Needs(R2)

Evaluation

Evaluation

temporal sensitivity (TS)

$$TS(i,a) = \mathrm{Var}(P(.|i,a))$$

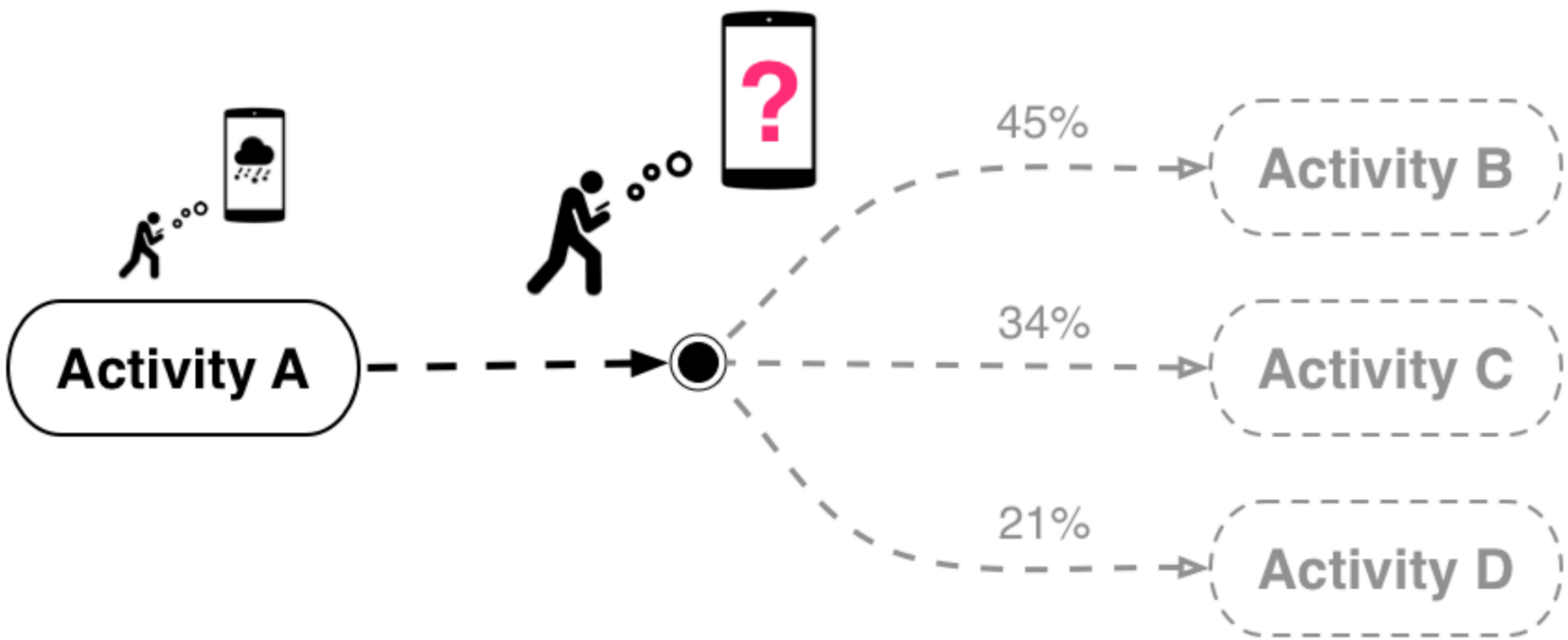


Outline

- Introduction
- (RQ1)Information Needs Related to Activities
- (RQ2)Analysis of Temporal Dynamics of Information Needs
- <u>(RQ3)Anticipating Information Needs</u>
- Conclusion

Anticipating Information Needs(R3)

Anticipation of user's future information needs given her last activity

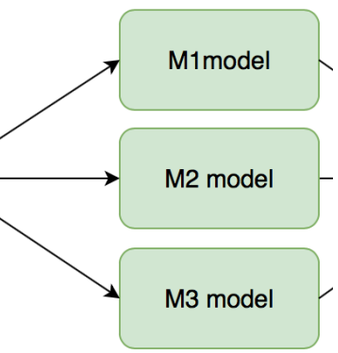


Anticipating Information Needs(R3)

transition probabilities

defined as activity session, where any two consecutive activities are separated by a maximum of 6 hours

$$P(a_j|a_i) = \frac{n(a_i \rightarrow a_j)}{\sum_k n(a_i \rightarrow a_k)}$$



Anticipating Information Needs(R3)

Proposed Models

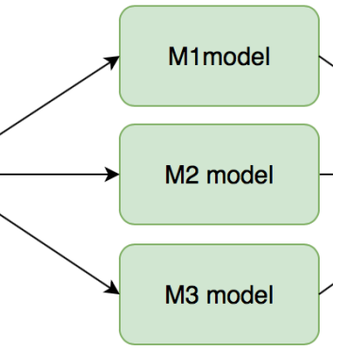
- M1 Consider all possible upcoming activity

$$P_{M1}(i|a_{last}) = \sum_{a_{next}} P(i|a_{next})P(a_{next}|a_{last})$$

- M2 Consider influence of the last activity

$$P_{M2}(i|a_{last}) = \gamma P(i|a_{last}) + (1-\gamma) \sum_{a_{next}} P(i|a_{next})P(a_{next}|a_{last})$$

$$\gamma = \frac{\sum_{i \in I, a \in A} P(t_{post}|i,a)}{|I| \cdot |A|}$$



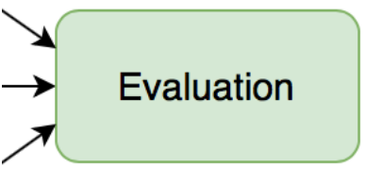
Anticipating Information Needs(R3)

Proposed Models

- M3 Consider the temporal dynamics of each information need individually

$$P_{M3}(i|a_{last}) \propto P(t_{post}|i, a_{last})P(i|a_{last}) + \sum_{a_{next}} P(t_{pre}|i, a_{next})P(i|a_{next})P(a_{next}|a_{last}).$$

Anticipating Information Needs(R3)

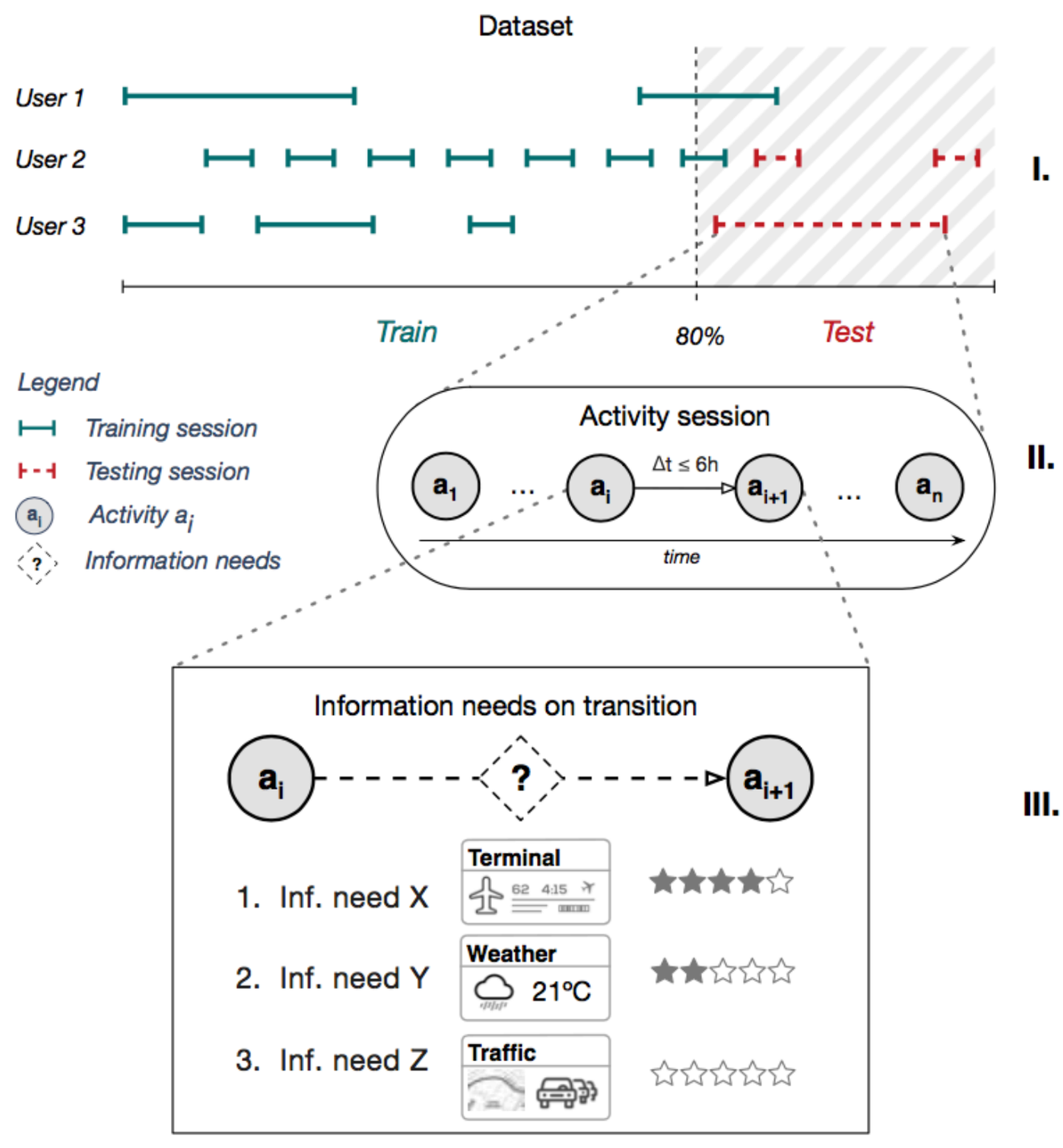


Evaluation

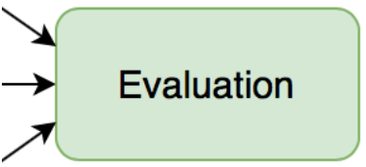
I.the check-in data are split into training and testing set

II.activity-to-activity transition probabilities

III.Crowd judges are tasked with evaluating the usefulness of individual information needs



Anticipating Information Needs(R3)



Evaluation

Table 6: Results for anticipating information needs, second-level activities. The -H suffix indicates the usage of hierarchical smoothing (only for second-level activities). Highest scores are boldfaced.

Model	top-level		second-level	
	NDCG@3	NDCG@5	NDCG@3	NDCG@5
M0	0.607	0.695	0.532	0.560
M1	0.824	0.828	0.712	0.705
M1-H			0.736	0.709
M2	**0.852**	**0.849**	**0.765**	**0.744**
M2-H			**0.765**	**0.744**
M3	0.756	0.780	0.735	0.741
M3-H			0.735	0.740

$$P_H(i|a^{l2}) = \lambda P(i|a^{l2}) + (1-\lambda)P(i|a^{l1}).$$

Table 7: Significance testing results (p-values).

Model	top-level		second-level	
	NDCG@3	NDCG@5	NDCG@3	NDCG@5
M0 vs. M1	0.0004	0.0068	0.0028	0.0014
M0 vs. M2	0.0002	0.0012	0.0009	0.0004
M0 vs. M3	0.0072	0.0131	0.0073	0.0007
M1 vs. M2	0.1183	0.0264	0.0307	0.0283
M1 vs. M3	0.0350	0.0541	0.8829	0.1345
M2 vs. M3	0.0021	0.0012	0.0199	0.9553

context-agnostic model M0, which always returns the most frequent information needs

Conclusion

A method for obtaining information needs and determining their relevance for various activities without relying directly on a large-scale search log

A detailed analysis of how the relevance of information needs changes over the course of an activity for different categories of activities

In a simulation experiment on historical check-ins combined with human judgments, we have shown that our models have good predictive performance