The Influence of Indoor Spatial Context on User Information Behaviours

Yongli Ren, Martin Tomko, Kevin Ong, Yuntian Brian Bai, Mark Sanderson

School of Computer Science and Information Technology

RMIT University

TRIIBE project

Feature	Value
Number of users:	120,548
Number of AP association:	907,084
Number of Web accesses:	18,008,018
Number of days covered:	406

Findings
- Basic Indoor Information Behaviours

- Distribution of the URLs over URL categories
  - 19% - Social Network
  - 15% - Computer and Internet Info
  - 13% - Content Delivery networks
  - 10% - Search Engines
  - 10% - Business and Economy

- Different from general mobile surfing
  - 3.2% for *Email and Social Network* in [4]
  - 23.1% for these two in our data set.

- Either the indoor context leads to a different information behaviour, or
- the information behaviour of mobile users has shifted since publication of [4].

Methodology

- We explore the associations between
  - users' physical spatial context, and
  - their Web information behaviours in the shopping mall.

- Spatial context
  - is investigated at the level of access point:
    - the spatial indoor context for each access point a_i is defined as a vector of shop categories $c_s^k \in C_s$.

$$E_i \quad \boxed{e_{i1}} \quad \boxed{e_{i2}} \quad \cdots \quad \boxed{e_{ik}} \quad \cdots \quad \boxed{e_{im}}$$

Number of shops located in the Voronoi cell of a_i and belong to $c_s^k \in C_s$

Methodology
- user information Behaviour

- The user information Behaviour at access point a_i is defined as a vector of Web page categories $c_w^k \in C_w$,

B_i b_{i1} b_{i2} . . . b_{ik} . . . b_{im}

The average number of URLs, which users issued through a_i and which belong to $c_w^k \in C_w$.

- At the level of access points,
  - the influence of spatial context on users' information behaviours
    - can be viewed as the correlation between B_i and B_j .
- In this study, we apply Pearson Correlation Coefficient (PCC)
  - between B_i and B_j
  - to investigate this association.

Findings
- Influence from Different Locations(1)

- There are differences in the types of shops served by different Wi-Fi APs.
- These shop categories describe the indoor context at each AP.

- Our hypothesis:
  - the proximity of different types of shops lead to a different Web Information behaviour.

- To investigate this,
  - we analyse the average PCC value for every pair of B_i .
  - the overall average PCC reflects the similarity of Web activates.
  - a small PCC value indicates
    - user information behaviour vary at different locations.

Findings
- Influence from Different Locations(2)

- When using all URL categories,
  - the PCC value is 0.9619,
    - which seems to show little differences among different APs.

- However, this is not true, and is caused by popular URL categories,
  - e.g. the top 5 URL categories take over 57.8% of overall URL records,
  - and this skewed Web behaviour introduces a bias in PCC calculation.
- Access entropy for URL category

$$H(c_w) = -\sum_{v \in S(c_w)} p(v \mid c_w) \log p(v \mid c_w)$$

$S(c_w)$ set of visits with access to c_w.

$p(v \mid c_w)$: the percentage of access to c_w during a visit v out of all visits (device per day)

- A high access entropy means that c_w is common among all users.

Findings
- Influence from Different Locations(3)

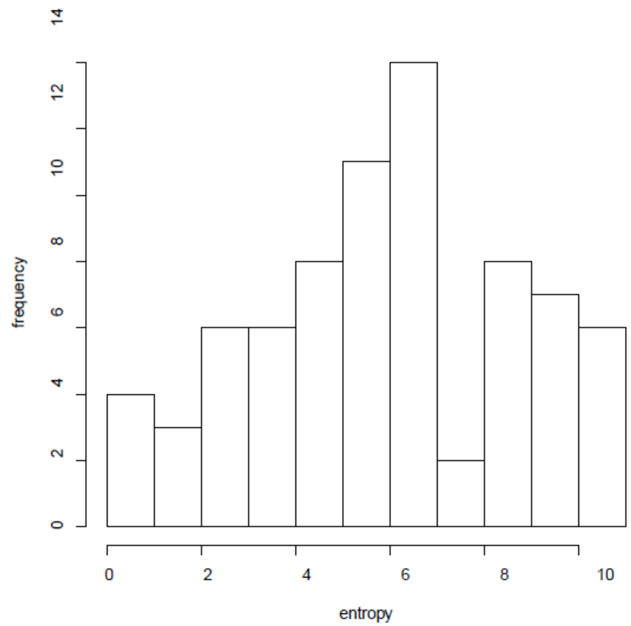

Fig. 1. The distribution of $H(c_w)$

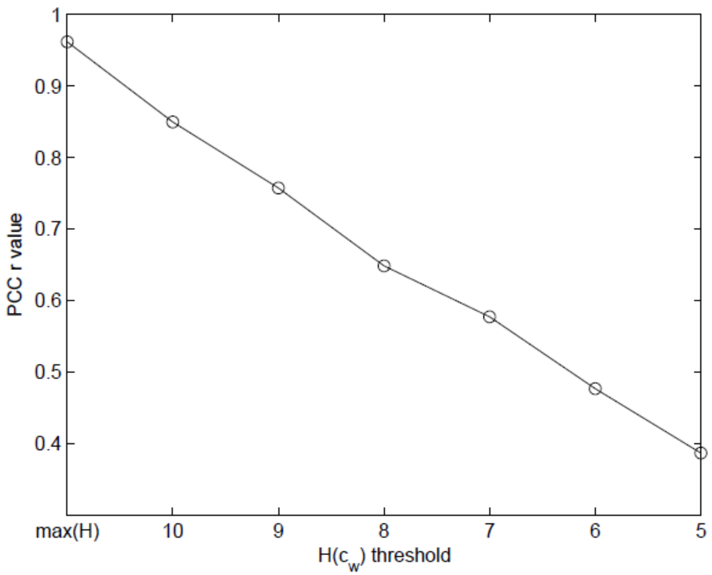

Fig. 2. PCC values without common c_w

When common URLs are removed, differences in information behaviours at different access points appear.

- $H(c_w)$ is defined over user visits.
- PCC is defined based on B_i at access point a_i.
- Thus, no logical influence between PCC and the removal of c_w.

Findings
- Influence of Indoor Context (1)

- To show the influence of indoor context, we apply
  - clustering algorithm to group similar access points based on E
- If users' information behaviour is influenced by the indoor context,
  - B_i in the same cluster should have higher PCC value (within), while
  - B_i in the different clusters should have lower PCC value (between).
- within: the average PCC of each pairs of B_i in the same cluster.

$$within = \frac{1}{k}\sum_{x=1}^{k}\left(\frac{2}{|t_x|(|t_x|-1)}\sum_{B_i \in t_x}\sum_{B_j \in t_x} PCC(B_i, B_j)\right)$$

- between: the average PCC of each pairs of B_i in different clusters.

$$between = \frac{1}{k}\sum_{x=1}^{k}\left(\frac{1}{|t_x|(|B|-1)}\sum_{B_i \in t_x}\sum_{B_j \notin t_x} PCC(B_i, B_j)\right)$$

- average: the average PCC of pairs of Bi.

$$average = \frac{1}{|B|(|B|-1)}\sum_{B_i}\sum_{B_j, i \neq j} PCC(B_i, B_j)$$

Findings
- Influence of Indoor Context (2)

	$H(c_w)$	PCC r value based on $\mathcal{B}$				
		k-means		random		average
		within	between	within	between	
Groups of Access Point based on $\mathcal{E}$	$H(c_w) \leqslant max(H(c_w))$	**0.9659**	0.9623	0.9609	0.9617	0.9619
	$H(c_w) \leqslant 10$	**0.8601**	0.8509	0.8493	0.8501	0.8498
	$H(c_w) \leqslant 9$	**0.7721**	0.7599	0.7564	0.7573	0.7573
	$H(c_w) \leqslant 8$	**0.6817**	0.6572	0.6493	0.6473	0.6483
	$H(c_w) \leqslant 7$	**0.6410**	0.5966	0.5767	0.5750	0.5770
	$H(c_w) \leqslant 6$	**0.5045**	0.4778	0.4755	0.4751	0.4763
	$H(c_w) \leqslant 5$	**0.4107**	0.3942	0.3821	0.3848	0.3863

Table 1. Correlation of user information behaviours in groups of access points with similar spatial context

Methods	t	p-value
within(k-means) VS *between*(k-means)	3.7962	0.0090
within(k-means) VS *within*(*random*)	3.5871	0.0115
within(k-means) VS *average*	3.4126	0.0143
within(*random*) VS *between*(*random*)	0.2526	0.8090
within(*random*) VS *average*	1.6007	0.1606

Table 2. Paired t-test results

Conclusion

- We found
  - The users' indoor information behaviour
    - manifests a significant location-based bias when the common information behaviour is excluded.
  - Users in similar indoor contexts
    - tend to access similar Web pages, while
  - users in dissimilar indoor contexts
    - tend to request dissimilar Web pages.

- This study has raised many new research questions:
  - What are the specific differences in user Web behaviours?
  - How to utilize the differences in information behaviours?

References

- [1] S. Algethami. Dubai Mall welcomes more than 200,000 shoppers a day. Gulfnews, 2014.
- [2] G. Biczok, S. Martinez, T. Jelle, and J. Krogstie. Navigating MazeMap: indoor human mobility, spatio-logical ties and future potential. CoRR, arXiv:1401, 2014.
- [3] K. Church, P. Ernest, and N. Oliver. Understanding Mobile Web and Mobile Search Use in Today's Dynamic Mobile Landscape. In MobileHCI'11, pages 67–76, 2011.
- [4] K. Church and B. Smyth. Understanding the intend behind mobile information needs. IUI, pages 247–256, 2009.
- [5] K. Church, B. Smyth, P. Cotter, and K. Bradley. Mobile information access: A study of emerging search behavior on the mobile Internet. ACM TWEB, 1(1), May 2007.
- [6] D. L. Davies and D. W. Bouldin. A Cluster Separation Measure. IEEE TPAMI, 1(2):224–227, 1979.
- [7] J. Teevan, A. Karlson, S. Amini, a. J. B. Brush, and J. Krumm. Understanding the importance of location, time, and people in mobile local search behavior. In MobileHCI '11, pages 77–80. ACM Press, 2011.
- [8] J. D. Vernor, M. F. Amundson, J. A. Johnson, and J. S. Rabianski. Shopping Center Appraisal and Analysis. 2009.
- [9] C. G. Wayne. A Better Space. Smithsonian Magazine, 2011.