Consistency of Digital Photo Classification Over Time

CONSISTENCY OF DIGITAL PHOTO CLASSIFICATION OVER TIME

Maria K Wolters[1], Elaine H Niven[2], Zeyu Wang[3], Robert H Logie[4]

1- School of Informatics, University of Edinburgh; 2 - Department of Psychology, University of Dundee;
3 - Department of Computer Science, Yale University; 4-Human Cognitive Neurosciences, University of Edinburgh, UK
Funded by the European Union under FP7 Grant 600826 ForgetIT.

Why Photo Categorisation?

Question: How to manage large amounts of personal and organisational data such that relevant information can be extracted from it easily and quickly?

ForgetIT Project Answer: leverage humans' ability to forget information that is either not relevant in the long term or not pertinent to the current context (Niederee et al. 2016).

Research question

Does the way in which participants sort their personal photos into groups (= evidence of categorisation) remain relatively constant over time? If yes, the categories reflected by these groups can be used to facilitate future annotation, storage, and retrieval.

Secondary aim: To inform the design of the personal information management component of the ForgetIT system (Maus et al., 2016)

Design

Quasi experimental design using sorting tasks are a standard way of highlighting similarities between items in Information Architecture (Tullis & Albert, 2013).

Hypotheses

Study 1: If there is a consistent, preferred categorisation pattern, then this should be reflected in strong similarity between Sorts S1, S21, and S3, and all three should be dissimilar from S22.

Study 2: We expect the number and size of event groups to be consistent across Sorts 2.1 and 2.2, given that the sequence of the events was the same.

Results

Study 1:
- Sort S1 and Sort S21 are similar, whereas Sort S1 and S22 and S21 and S22 are dissimilar ($p<0.001$; cf. Table 1)
- No significant differences between delays ($p=0.33$).
- Sort S1 and Sort S3 (year-recall) are relatively similar only for people who returned after a day (M: 0.66, SD 0.2), but not for the other groups (M: 0.32, week delay; M: 0.33, month delay; cf. Figure 1)

Study 2:
- Both event groups and non-event groups of similar size and number both at Time 1 and a month later, at Time 2. (c.f. Table 2)

Discussion

- Categorisation of photos relatively stable over time delays of up to a month.
- Stable categorisation strategies are highly idiosyncratic, as expected from Human Computer Interaction research on photo work (e.g. Kirk, Sellen, Rother, & Wood, 2006)
- After longer delays such as a year, categorisations may change, especially if they were not rehearsed soon after the first categorisation was made.

Future Work

- Categorisation changes over a year to be tested in specific longitudinal study.
- Card sorting design to be complemented with qualitative work

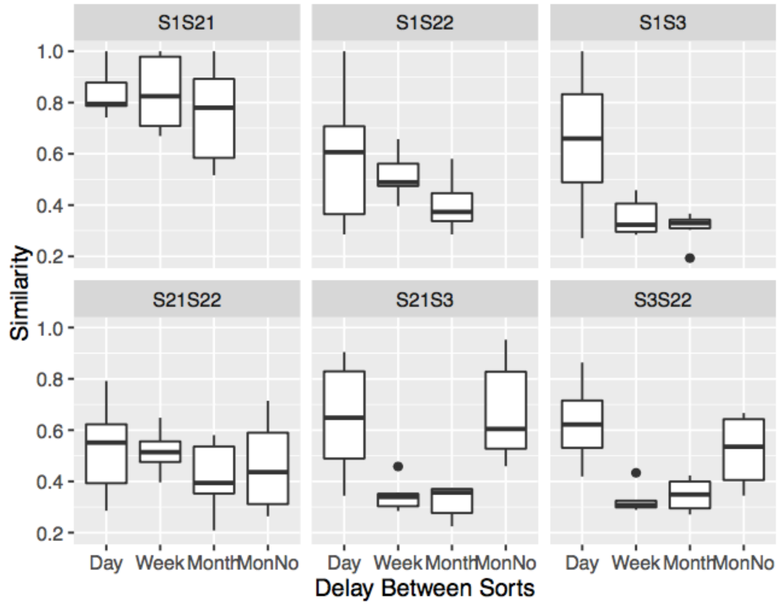

Figure 1: Group Similarities between Time 1, Time 2 Sort 1 (S21), Time 2 Sort 2 (S22), and Time 3 for all four participant groups. Day, Week, Month: Recall plus Sort on the day, MonthNo: No sort on the day, Recall after a month

The Data Set: Festival Studies

Within the ForgeitIT project, the University of Edinburgh team conducted two studies that required participants to document their experience of the Edinburgh Festival Fringe using digital photos (Niven et al. 2014, 2015). As part of these studies, participants were asked to sort their photos into groups.

Study 1: An Hour on the Royal Mile
Participants spent an hour at the Royal Mile street festival, taking photos every three minutes. They were debriefed immediately afterwards. Participants returned after a day (n=20), a week (n=18), and a month (n=36).
All except 18 people in the month group (referred to as month no-sort) sorted their photos into groups at time 1 (***Sort S1***). On returning, people were asked to sort their photos twice (***Sort S21*** and ***Sort S22***), and to make ***S22*** as different as possible from ***S21***.
6 participants per group then returned 11 months later and performed the sorting task again (***Sort 3***).

Similarity between groups of photos was calculated using the Jaccard index (for a detailed description of the method, see Logie et al., 2016). Results are normalised so that 0 = maximum dissimilarity, 1 = maximum similarity.

Study 2: A Day at the Festival
Participants (n=22, 21 returned) documented a day at the Festival Fringe, taking 40-80 photos of their experience. All participants were debriefed the next day, and returned a month later to annotate their photos using the ForgetIT PIMO.
At both debriefings, they were asked to sort their photos into groups twice, once according to events and once according to criteria of their own choosing. The sorts from debriefing 1 will be referred to as ***Sort 2.1*** and that from debriefing 2 as ***Sort 2.2***.

In both studies, participants were asked to name or describe the groups generated.

Statistical Analysis

The statistical significance of the difference between two sorts regarding a specific characteristic was assessed using an asymptotic Wilcoxon Mann-Whitney test (R package coin). Differences between groups of participants in Study 1 were assessed using the Kruskal-Wallis test.

	n	Sort S1/S21	Sort S1/S22	Sort S21/S22
All	56	0.75 (SD: 0.2)	0.49 (SD: 0.2)	0.47 (SD: 0.1)
Day	20	0.80 (SD: 0.2)	0.53 (SD: 0.3)	0.49 (SD: 0.2)
Week	18	0.77 (SD: 0.2)	0.50 (SD: 0.2)	0.49 (SD: 0.1)
Month	18	0.68 (SD: 0.2)	0.43 (SD: 0.2)	0.40 (SD: 0.1)

Table 1: Mean similarity between sorts at Time 1 and Time 2 in Study 1

		Sort 2.1	Sort 2.2	Sig.
Events	# groups size	5.5 (range: 2-8) M: 11, SD: 6	5 (range: 2-11) M: 12, SD: 7.5	p=0.6 p=0.8
Non-Event Categories	# groups size	4 (range: 2-8) M: 16 (SD: 8)	3 (range: 2-7) M: 15 (SD: 7)	p=0.9 p=0.7

Table 2: Number (median) and size (mean) of groups in Study 2

References and further information:

http://mariawolters.net/psychonomics2016

maria.wolters@ed.ac.uk for the team