Aspect-Based Personalized Text Summarization

Shlomo Berkovsky[1], Timothy Baldwin[1], and Ingrid Zukerman[2]

[1] University of Melbourne, Australia
{shlomo,tim}@csse.unimelb.edu.au
[2] Monash University, Australia
ingrid@csse.monash.edu.au

Abstract. This work investigates user attitudes towards personalized summaries generated from a coarse-grained user model based on document aspects. We explore user preferences for summaries at differing degrees of fit with their stated interests, the impact of length on user ratings, and the faithfulness of personalized and general summaries.

1 Introduction

Exponential growth in information availability has increased the need for intelligent filtering and efficient presentation methods. *Personalized* summarization [1] presents users with document extracts that are of interest to them, as defined by a user model [2] (c.f. general summarization, which is oblivious to user interests).

This research is novel in evaluating user attitudes towards personalized summarization of aspect-based documents, i.e., documents that can be partitioned into mutually-exclusive sub-documents that relate to different subject areas. In contrast, past research on personalized summarization has relied on an intrinsic representation of user interests via keywords or document categories [1].

Our results show that: (1) users prefer personalized summaries which accurately reflect their interests, supporting the findings of [1]; (2) users have a preferred summary length, and disprefer over-long or over-short summaries; and (3) users perceive the faithfulness to the original document of personalized and general summaries to be roughly equivalent.

2 Data Representation and Personalized Summarization

The domain for this research is natural science, to fit in with the scope of the Kubadji project (http://www.kubadji.org), which is focused on personalization in museums. For example, consider the following document d about blue whales, extracted and pre-processed from a longer Wikipedia article into coherent logical aspects as follows.

d_1	*The blue whale is a marine mammal belonging to the family of baleen whales. This family also includes the Humpback, Fin, and Minke Whales. Due to its yellow underparts, the blue whale is often called the sulphur-bottom.*

W. Nejdl et al. (Eds.): AH 2008, LNCS 5149, pp. 267–270, 2008.

d_2	*Blue whales are believed to be the largest animals to have ever lived. They reach 33 meters in length and 200 tonnes in weight. When breathing, they emit a spectacular vertical column blow of up to 12 meters.*
d_3	*The London Natural History Museum contains a life-size model of a blue whale. Living whales may be encountered in Saint Lawrence Gulf. It was represented as symbol of size and strength in the movie Doctor Dolittle.*

Each of the sub-documents d_1, d_2, d_3 represents a different aspect or subject area, viz biological taxonomy, physical dimensions and popular culture, respectively. Assuming a relatively homogeneous document collection (as is the case with curated data) and a coarse-grained set of aspects, we can expect to be able to partition other documents about marine animals according to a single set of aspects (we considered the above three aspects, plus reproduction and life and threats and dangers).

The same set of aspects was also used as the basis of a content-based user model [3], where a user's interests are represented by a vector of domain aspects. Our representation was based on the following 4-point scale of interest in aspects: 0=*no interest*, 1=*low*, 2=*moderate*, and 3=*high*. For example, a user with moderate interest in biological taxonomy, high interest in physical dimensions and low interest in popular culture would be represented by $UM = \{LI_i\} = \{2, 3, 1\}$, where LI_i denotes the level of interest in aspect i.

The aspect-based representation of the user models facilitates the generation of personalized summaries, where the amount of text for a given aspect is proportional to the user's interest in it. For our experiments, we prepared a ranked list of n sentences for each aspect i, and included in the summary the first m sentences for a given aspect, where $\frac{m}{n}$ is proportional to LI_i. For example, a personalized summary of d based on the above model $UM = \{2, 3, 1\}$ is:

The blue whale is a marine mammal belonging to the family of baleen whales. This family also includes the Humpback, Fin, and Minke Whales.
Blue whales are believed to be the largest animals to have ever lived. They reach 33 meters in length and 200 tonnes in weight. When breathing, they emit a spectacular vertical column blow of up to 12 meters.
The London Natural History Museum contains a life-size model of a blue whale.

3 User Study

We conducted three experiments to assess different aspects of users' attitudes towards personalized document summarization.

Experiment 1 evaluated whether the personalization of summaries has the desired effect, i.e., whether personalized summaries adjusted to actual user interests are preferable to those adjusted to other interests. Four summaries were composed for each of four documents, each of which contained the above mentioned five aspects. Each summary was adjusted to one of 16 pre-determined pseudo-models UM_{ps}. These pseudo-models were derived using fractional factorial design [4], such that they uniformly cover the search space of possible pseudo-models (4^5 states). A total of 19 users provided a rating *eval* (from 1=*bad* to 5=*good*) for each of the 16 summaries. After rating the 16 summaries, the users were asked to explicitly provide ratings for their interest in each of the five

Table 1. Average user evaluation at differing levels of user model fit

$sim(UM_r, UM_{ps})$	$(-1, -\frac{2}{3})$	$[-\frac{2}{3}, -\frac{1}{3})$	$[-\frac{1}{3}, 0)$	$[0, \frac{1}{3})$	$[\frac{1}{3}, \frac{2}{3})$	$[\frac{2}{3}, 1)$
$\overline{eval}$	2.29	2.59	2.63	2.77	2.85	3.11

aspects. This was taken to be the real user model UM_r, acknowledging that a user's self-perception may differ from actuality.

Given the user model UM_r and each pseudo-model UM_{ps}, we calculated their similarity $sim(UM_r, UM_{ps})$ using Pearson's Correlation. This allows us to measure the relative fit between the two models, and hence analyze the correlation between the rating *eval* of the summaries and the faithfulness of the personalization to the actual user interests. In this analysis, we discretized the similarity values into six equal-width bins over the range of the Pearson Correlation Coefficient $[-1, 1]$, and calculated the average user rating $\overline{eval}$ in each bin.

Table 1 shows the average user rating ($\overline{eval}$) at each level of fit between the pseudo-model and the real user model ($sim(UM_r, UM_{ps})$).[1] The ratings of personalized summaries increase monotonically as the level of fit increases. This demonstrates that, as expected, users preferred summaries matching their actual interests. This finding was separately validated via a linear regression analysis of the ratings at the different levels of Pearson's Correlation (without discretization), which returned a right-increasing function.

Experiment 2 assessed the impact of compression on the ratings given by users. This experiment was conducted after the first experiment, i.e., after eliciting the real user model UM_r. We generated three personalized summaries at different compression levels: (1) an original-length summary adjusted to $UM_r = \{LI_i\}$, (2) a *lengthened* summary adjusted to $UM_l = \{\alpha LI_i\}$, and (3) a *shortened* summary adjusted to $UM_s = \{(1/\alpha)LI_i\}$; α was set to 1.5.

19 users were shown three randomly-ordered summaries (at the three levels of compression) for each of two previously unseen documents, and were asked to rate each summary. We obtained a total of 114 ratings — 38 for each type of summary. The average rating $\overline{eval}$ was 3.32 for the original length, 2.95 for the lengthened, and 2.37 for the shortened summaries (all differences statistically significant: $p = 2.0 \times 10^{-2}$ for lengthened and $p = 9.3 \times 10^{-7}$ for shortened). This shows that users disliked personalized summaries that were too long or too short, although they were less averse to overly-long summaries.[2]

Experiment 3 evaluated the perceived faithfulness of the personalized summaries to the original documents. We generated two summaries: (1) a personalized summary adjusted to UM_r — the user model elicited in the first experiment; and (2) a general summary adjusted to a model with equal interest levels in all aspects.

[1] These results do not include the ratings for 4 users with a uniform user model UM_r, due to a divide-by-zero error for Pearson's correlation.

[2] Noting that users were primed for summary length in the first experiment, where the average summary length was 12.5 sentences.

19 users were shown two original (previously unseen) documents, and a general and personalised summary for each. They were asked to rate the faithfulness *rel* of the summaries to the original document ($1 \leq rel \leq 5$). We obtained a total of 76 ratings — 38 for each type of summary. The average faithfulness $\overline{rel}$ was 3.11 for the personalized and 3.21 for the general summaries.[3] Although the faithfulness of the general summaries was slightly higher, the results were not statistically significant, i.e., the two types of summaries are comparable in terms of faithfulness to the original document.

4 Conclusions and Future Research

We have conducted three studies to evaluate users' attitudes towards aspect-based, personalized document summaries. The results of our studies show that the better the fit between the real user model and the user model on which a summary is based, the higher the user's rating for this summary; and that there is a preferred length for personalized summaries. Evaluating the perceived faithfulness of a summary to the original document did not show a significant difference between personalized and general summaries. This leads to the conclusion that personalized summaries are both appropriate and liked by users.

This conclusion motivates further research in the Kubadji project, where we intend to harness user models of museum visitors [5] to dynamically generate personalized exhibit summaries.

Acknowledgments

This research was supported in part by Discovery grant DP0770931 from the Australian Research Council. We would like to thank Willy for help in setting up the web experiment.

References

1. Díaz, A., Gervás, P.: User-model based personalized summarization. Information Processing and Management 43(6), 1715–1734 (2007)
2. Brusilovsky, P., Millán, E.: User models for adaptive hypermedia and adaptive educational systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) The Adaptive Web: Methods and Strategies of Web Personalization, pp. 3–53. Springer, Heidelberg (2007)
3. Pazzani, M.J., Billsus, D.: Content-based recommendation systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) The Adaptive Web: Methods and Strategies of Web Personalization, pp. 325–341. Springer, Heidelberg (2007)
4. Box, G.E.P., Hunter, J.S., Hunter, W.G.: Statistics for Experimenters: Design, Innovation, and Discovery. Wiley, Chichester (2005)
5. Bohnert, F., Zukerman, I., Berkovsky, S., Baldwin, T., Sonenberg, L.: Using collaborative models to adaptively predict visitor locations in museums. In: Proceedings of Adaptive Hypermedia 2008, Hannover, Germany (2008)

[3] The average sentence overlap between the personalized and general summaries was high at 0.83.