

Distribution of User-perceived Usefulness of Four Presentation Styles of Opinion Summarization

Ning Sa¹, Xiaojun Yuan¹, Grace Begany¹, Huahai Yang²

¹College of Engineering and Applied Sciences, University at Albany, State University of New York, USA

²Juji, Inc., Saratoga, USA

Abstract

Four opinion summarization styles, including tag clouds, aspect oriented sentiments, group samples, and paragraph summaries were compared in a card sorting lab environment. Thirty four participants sorted thirty two cards and grouped them into five usefulness categories (each presentation style has eight cards). We found that the users spent the shortest time on cards in “not at all useful” category. Participants spent shorter time on “extremely useful” cards than on “somewhat useful”, “useful”, and “very useful” cards. It seems that Tag clouds and Aspect oriented sentiments needed less time to view than the other two styles. They are the major styles in “not at all useful” and “extremely useful” categories. Paragraph summaries and Group samples requested more time than the other styles and they occupy about 50% in “somewhat useful”, “useful”, and “very useful”. The findings were discussed and compared with previous results.

Keywords: Opinion summarization; usability; presentation styles

doi: 10.9776/16576

Copyright: Copyright is held by the authors.

Acknowledgements: we thank all the participants in our experiment

Contact: nsa@albany.edu, xyuan@albany.edu, gbegany@albany.edu, hyang@juji-inc.com

1 Introduction

Online customers rely on other users' review in purchasing product. However, sometimes the amount of the online customer reviews is huge and most of them are not relevant to what the customer really wants. A summary of the opinions or reviews will to a large extent benefit the customers. The opinion summary can be organized in various presentation styles. The existing evaluations of different styles of opinion summarization have focused on effectiveness (e.g. precise and recall) or usability of a specific presentation style. In our research project, we aimed at comparing the usability of four different presentation styles from the user perspectives: Tag clouds (TAG), Aspect oriented sentiments (ASP), Paragraph summaries (PRG), and Group samples (GRP). In this paper, we report the relationship of the presentation styles, the user perceived usefulness, and the time of viewing each opinion summary in each usefulness category.

2 Related Work

The style of Tag clouds contains words and phrases which are most common in the opinions. It is the most popular summarization style on internet today (Brooks & Montanez, 2006). Different font size, weight, and color are often used to indicate the frequency of the words. Bateman, Gutwin, and Nacenta (2008) investigated the effectiveness of different visual features. When manipulated separately, font size and font weight were found to be more “visually important”. Aspect oriented sentiments contain important aspects or topics extracted from the opinions by using text mining (Lu, Zhai, & Sundaresan, 2009). Samples of original opinions are categorized into “positive”, “neutral”, and “negative”. The statistics of each category are also presented to the users. Lerman and McDonald (2009) evaluated three different presentation styles composed of Sentiment Match and Aspect Coverage and found that sentiment-informed summaries were preferred by the users. Paragraph summaries are short text passages generated from automatic text summarization systems (Hovy & Lin, 1998). In the work of Carenini, Cheung, and Pauls (2013), natural language summaries of online customer electronics reviews were presented in an interactive multimedia interface. The users reported that the summaries were informative and the interface was satisfying. In group samples, customer opinions are clustered into groups and the representatives of each group are presented to the users (Ando, Boguraev, Byrd, & Neff, 2000). The grouping and classification methods have been used in Information Retrieval and the effectiveness has been demonstrated (Leuski, 2001). An example of each of the four styles is displayed in Figure 1.

As mentioned above, there exist various summarization styles. However, previous studies have focused on the effectiveness and usability of one specific presentation style. The usability comparison of the different styles is missing. We aimed at filling this gap in our research project. In (Yuan, Sa, Begany, & Yang, 2015), factors affecting user preference of presentation styles were investigated by conducting content analysis. It was found that “comprehensiveness”, “Organization/Categorization”, and “Ease of Use” were the main reasons why the users preferred one style over another. In (Yuan, Sa, Begany & Yang, 2016), we analyzed user rated usefulness of the presentation styles and found that ASP was evaluated significantly more useful than the other three presentation styles. The users also spent significantly less time on ASP and TAG than on PRG and GRP. In this paper, we report the relationship between user-perceived usefulness and the time to read the summary in each usefulness category. The components of each usefulness category are also studied.

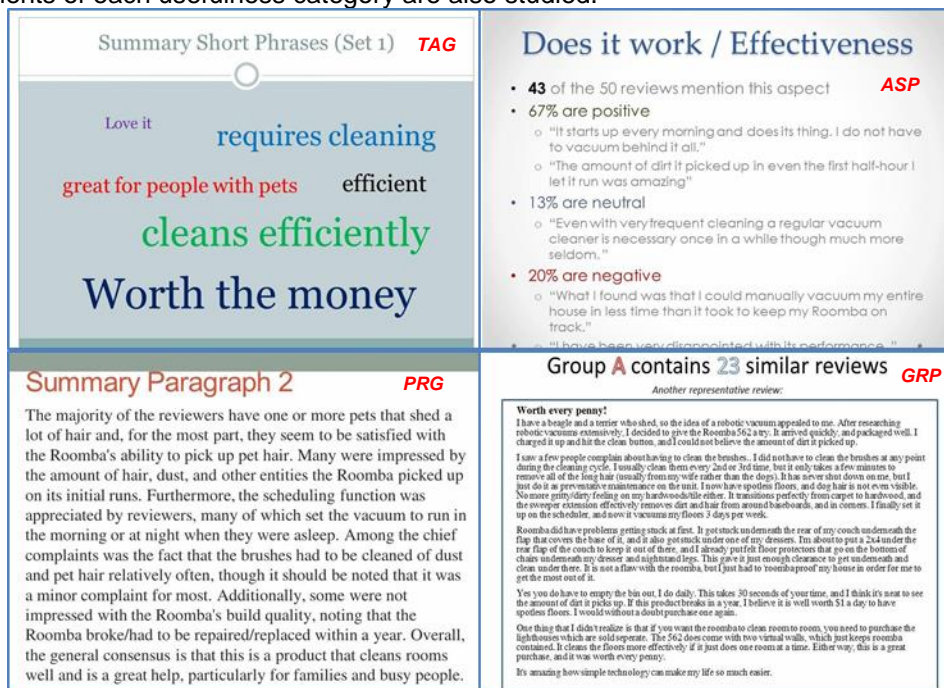


Figure 1. Example of Four Presentation Styles in Opinion Summarization: TAG (upper left), ASP (upper right), PRG (lower left), and GRP (lower right)

3 Methodology

The research project was composed of two phases. In the first phase, opinion summarizations on iRoomba cleaning robot were generated and grouped into four different styles by using Amazon Turkers. The details about this phase, including how the summaries were generated and how the quality of the content was controlled, could be found from (Yuan et al., 2015). In the second phase, the generated summarizations were presented to the users to read and evaluate in a card sorting lab environment. Thirty-four participants were recruited from a US university.

3.1 Experimental design

To compare the four presentation styles, a within-subject card-sorting experiment was designed and conducted. Thirty two cards were prepared with eight cards for each style. Figure 1 shows example cards of the four presentation styles. The cards were randomly organized and presented to the participants. All the participants read all the 32 cards. After reading each card, the participant needed to drag the card and put it into one of the five usefulness categories: “not at all useful”, “somewhat useful”, “useful”, “very useful”, and “extremely useful”.

At the beginning of the experiment, the participants filled the consent form followed by an entry questionnaire. Then they performed the card-sorting task. They were encouraged to think aloud during the experiment. After completing the task, the participants were asked about opinions on the four presentation styles and their thought on the experiment. Morae (<https://www.techsmith.com/morae.html>) software was used to record the whole process and the desktop activities.

4 Results and Discussion

Figure 2 shows the relationship between the time spent on a card and the user-perceived usefulness. A one-way ANOVA analysis indicated that the time participants spent viewing the cards significantly depended on the usefulness of the cards ($F = 4.18$ (4; 1083), $p = .002$). A Tukey post-hoc test revealed that the average time spent on ‘not at all useful’ cards (Mean=50.41, SD=40.42) was significantly shorter than that on ‘useful’ cards (Mean=69.37, SD=75.41) and ‘very useful’ cards (Mean=74.25, SD=69.08). Also time used to view ‘very useful’ cards was significantly longer than the time used to view ‘extremely useful’ cards (Mean=55.42, SD=42.68). As Figure 2 shows, cards categorized as “not at all useful” were where users spent the least time. The time to view the card increased on “somewhat useful,” “useful,” and “very useful” cards, but decreased on “extremely useful” cards.

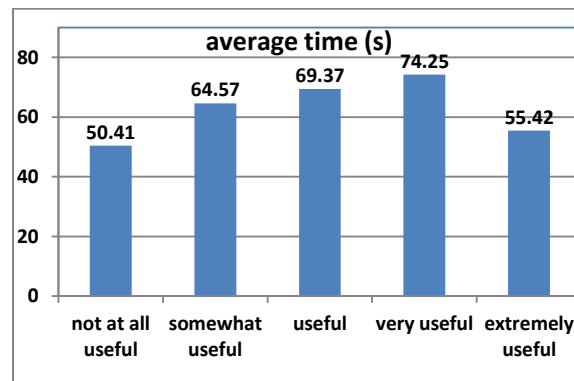


Figure 1. Average time to complete the task (in seconds) in each usefulness category

Table 1 shows the proportion of the four styles in each usefulness category. TAG is the majority type of card in “not at all useful” and “somewhat useful”. GRP and PRG take the largest proportion in “useful” and “very useful”, respectively though the difference is not obvious. In “extremely useful” group, ASP accounts for nearly 40% of the cards.

	ASP	GRP	PRG	TAG
not at all useful	18.71%	17.42%	22.58%	<u>41.29%</u>
somewhat useful	17.74%	27.42%	23.79%	<u>31.05%</u>
useful	23.98%	<u>26.42%</u>	24.39%	25.20%
very useful	25.00%	28.70%	<u>31.94%</u>	14.35%
extremely useful	<u>38.57%</u>	22.42%	21.97%	17.04%

Table 1. Proportion of each presentation style in different usefulness categories

From (Yuan et al., 2016), it was found that the average time to view TAG cards and ASP cards were significantly less than that on GRP and PRG cards. In “not at all useful” and “extremely useful” categories, TAG and ASP are the two major styles (nearly 60%). On the other hand, GRP and PRG, which took longer time to complete, comprised the majority (>50%) of the three usefulness categories in the mid-range (“somewhat useful” to “very useful”). Correspondingly, the average time spent on “not at all useful” and “extremely useful” is less than that on “somewhat useful”, “useful”, and “very useful”, as displayed in Figure 2. These findings are consistent with our previous results. In (Yuan et al., 2015) PRG was found to be the most informative but was also “too long to read” and sometimes “not organized”. GRP received similar comments. ASP was found to be easy to use, “clear, brief”, and comprehensive. TAG was “very fast” but “wouldn’t drive my decision making”.

5 Conclusion

In this paper, we report the relationship between user-perceived usefulness and time to view a card in each usefulness category. We found that the users spent less time viewing cards in “not at all useful” category, where TAG was the majority type of cards, and “extremely useful” category whose major component was ASP. These findings are consistent with our previous results (Yuan et al., 2015) where TAG was reported to be fast but least useful and ASP was brief and most useful.

6 References

- Ando, R. K., Boguraev, B. K., Byrd, R. J., & Neff, M. S. (2000). Multi-document summarization by visualizing topical content. In *Proceedings of the 2000 NAACL-ANLP Workshop on Automatic Summarization* (pp. 79–98). Stroudsburg, PA, USA: Association for Computational Linguistics. Retrieved from <http://dl.acm.org/citation.cfm?id=1567564.1567573>
- Bateman, S., Gutwin, C., & Nacenta, M. (2008). Seeing things in the clouds: The effect of visual features on tag cloud selections. In *Proceedings of the Nineteenth ACM Conference on Hypertext and Hypermedia* (pp. 193–202). New York, NY, USA: ACM. <http://doi.org/10.1145/1379092.1379130>
- Brooks, C. H., & Montanez, N. (2006). Improved annotation of the blogosphere via autotagging and hierarchical clustering. In *Proceedings of the 15th International Conference on World Wide Web* (pp. 625–632). New York, NY, USA: ACM. <http://doi.org/10.1145/1135777.1135869>
- Carenini, G., Cheung, J. C. K., & Pauls, A. (2013). Multi-document summarization of evaluative text. *Computational Intelligence*, 29(4), 545–576. <http://doi.org/10.1111/j.1467-8640.2012.00417.x>
- Hovy, E., & Lin, C.-Y. (1998). Automated text summarization and the SUMMARIST system. In *Proceedings of a Workshop on Held at Baltimore, Maryland: October 13-15, 1998* (pp. 197–214). Stroudsburg, PA, USA: Association for Computational Linguistics. <http://doi.org/10.3115/1119089.1119121>
- Lerman, K., & McDonald, R. (2009). Contrastive summarization: An experiment with consumer reviews. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers* (pp. 113–116). Stroudsburg, PA, USA: Association for Computational Linguistics. Retrieved from <http://dl.acm.org/citation.cfm?id=1620853.1620886>
- Leuski, A. (2001). Evaluating document clustering for interactive information retrieval. In *Proceedings of the Tenth International Conference on Information and Knowledge Management* (pp. 33–40). New York, NY, USA: ACM. <http://doi.org/10.1145/502585.502592>
- Lu, Y., Zhai, C., & Sundaresan, N. (2009). Rated aspect summarization of short comments. In *Proceedings of the 18th International Conference on World Wide Web* (pp. 131–140). New York, NY, USA: ACM. <http://doi.org/10.1145/1526709.1526728>
- Yuan, X., Sa, N., Begany, G., & Yang, H. (2015). What users prefer and why: A user study on effective presentation styles of opinion summarization. In J. Abascal, S. Barbosa, M. Fetter, T. Gross, P. Palanque, & M. Winckler (Eds.), *Human-Computer Interaction – INTERACT 2015* (pp. 249–264). Springer International Publishing. Retrieved from http://link.springer.com/chapter/10.1007/978-3-319-22668-2_20
- Yuan, X., Sa, N., Begany, G., & Yang, H. (2016). Presentation styles on user perceived usefulness in opinion summarization. To be presented in *Hawaii International Conference on System Sciences (HICSS) 2016*