

Polarized Review Summarization as Decision Making Tool

Paolo Cremonesi, Raffaele Facendola, Franca Garzotto,
Matteo Guarnerio, Mattia Natali, Roberto Pagano
Politecnico di Milano, DEIB
P.zza Leonardo da Vinci, 32
Milan, Italy
{first name.last name}@polimi.it

ABSTRACT

When choosing an hotel, a restaurant or a movie, many people rely on the reviews available on the Web. However, this huge amount of opinions make it difficult for users to have a comprehensive vision of the crowd judgments and to make an optimal decision. In this work we provide evidence that automatic text summarization of reviews can be used to design Web applications able to effectively reduce the decision making effort in domains where decisions are based upon the opinion of the crowd.

1. INTRODUCTION

People's decision making is ever more influenced by the "voice of the crowd". When choosing an hotel, a restaurant or a movie many people rely on the reviews available on the Web, trusting other users' opinions more than the "official" ones by critics, guidebooks, experts, or similar. At the same time, we are witnessing a proliferation of user generated content: TripAdvisor, for instance, has recently passed the 150 million mark in terms of reviews and opinions posted to its site, now collecting more than 90 user contributions per minute ¹.

This huge mass of web based user judgements encapsulate a potentially reliable "ground truth" that, in principle, can drive the decision making process. Yet, too many pieces of information make it difficult for decision makers to get a comprehensive vision of the crowd judgement and its global polarity. The goal of our research is to help users overcome this problem. Bounded Rationality Theory [1] posits that decision-makers continuously try to find a balance between the knowledge needed for the optimal decision and the effort required to process such knowledge. When they lack the ability and resources to arrive at the optimal solution,

The work leading to these results has received partial funding from the European Union's Seventh Frame work Programme (FP7/2007-2013) under CrowdRec Grant Agreement number 610594

¹<http://ir.tripadvisor.com/releasedetail.cfm?ReleaseID=827994>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author. Copyright is held by the owner/author(s).
AVI '14, May 27 - 30, 2014, Como, Italy
ACM 978-1-4503-2775-6/14/05.
<http://dx.doi.org/10.1145/2598153.2600047>

they tend towards a suboptimal result after having greatly simplified the choices available. In line with this theory, our approach is to decrease user's effort by reducing the amount of information to be cognitively processed, offering users only a distilled vision of the voice of the crowd. More precisely our approach is to present summaries of all available reviews that capture the salient aspects of the crowd's judgement and are appropriate to build at least suboptimal solutions.

Summarization is emerging as a topic of growing importance. Summly², a news summary service designed to help simplify the way users consume news on mobile devices, was recently acquired by Yahoo for 30 million dollars. In comparison with this service, as well as with other academic works ([6, 5]) we have a different motivation: supporting decision making. In addition, we employ a different and novel technique that does not summarize a *single* text but *many* texts. Finally we provide two summaries, respectively summarizing positive and negative judgements, in order to offer a potentially more balanced and trustable vision of crowd's opinion.

We have performed our studies in the domain of *restaurant booking*. We conducted two separate user experiments: the first to choose the best summarization technique for the target domain, the second to demonstrate the effectiveness of the summarization technique as a tool to reduce the decision making effort.

2. SUMMARIZATION TECHNIQUE

The first problem we are addressing is to find the "best" summarization technique. Since we are interested to provide the users with the distilled user-based ground truth on items offering a balanced view of positive and negative opinions, we partition the reviews in a *positive* and a *negative* set.

The polarity of a review is derived by the user's rating: reviews with a rating greater than 3 (on a 5 star rating scale) are considered as positive, while reviews with a rating smaller than 3 are considered as negative.

For each item (i.e., restaurant) we apply a summarization algorithm to both the positive and negative sets of reviews, thus producing two distinct summaries. We use Sumy³, a framework that embeds 5 different summarization techniques: LSA [4], TextRank [8], LexRank [3], Luhn [7], Edmunson [2]. The algorithms generate the summaries by extracting sentences from the actual reviews without altering them. We also developed a technique that extract sentences

²<http://summly.com/>

³<https://github.com/miso-belica/sumy>

at random for control purposes. We limit the length of the positive and negative summaries to 7 sentences maximum.

The research question we wish to answer is to assess which technique among the proposed ones generates the summary which better represents the entire set of reviews.

To answer this question, we performed an online study with 33 participants. For each participant: (i) a random restaurant is picked among the entire dataset; (ii) the restaurant’s reviews are provided to the user alongside with the 12 summaries (two for each of the six techniques, one for the positive opinions and one for the negative ones); (iii) the user is asked to rank the summaries based on his opinion on how well the summary summarizes the reviews.

The best technique according to this study is Edmunson [2] which uses a statistical approach with a different weighting scheme for different classes of features (cue words, keywords, title, location of words inside the corpus).

3. SUMMARIES AND DECISION MAKING

In the second user experiment we studied if the polarized summarization technique described in Section 2 provides benefits for the user decision making process.

For this purpose we have developed a web application that mimics the functionality of TripAdvisor, but limited to the restaurants in New York City.

Each user involved in the experiment was asked to simulate the booking of a restaurant. Users were randomly split into two experimental conditions: (i) with original TripAdvisor reviews (Figure 1a) and (ii) with positive and negative summaries only (Figure 1b). A total of 108 users participated to this study.

In order to estimate the decision making effort, we measured the the average time each user spent in reading the description pages of restaurants. Users in the second experimental condition – with summaries of positive and negative reviews for each restaurant – spent, on average, *half of the time* with respect to the users in the first experimental condition – with the full set of reviews.

4. DISCUSSION AND CONCLUSIONS

The results from the second user experiment confirm that the summarization of reviews effectively reduce the effort of users in the decision making process.

Although this is only a preliminary study, its internal validity is supported by the accuracy of the research design. In terms of external validity, the applicability of our results might not be confined to the specific domain of restaurant booking, as in many decision making domains users rely their decision on the opinion of the crowd.

In order to provide better evidence of the effectiveness of summarization as a decision making tool we need to strengthen our study by (i) providing a better engagement for users, (ii) by collecting data from a larger user base and (iii) by collecting other metrics to measure the effectiveness of the decision making process.

5. REFERENCES

[1] P. Cremonesi, F. Garzotto, and R. Turrin. Investigating the persuasion potential of recommender systems from a quality perspective: An empirical study. *ACM Trans. Interact. Intell. Syst.*, 2(2):11:1–11:41, June 2012.

[2] H. P. Edmundson. New methods in automatic extracting. *Journal of the ACM (JACM)*, 16(2):264–285, 1969.

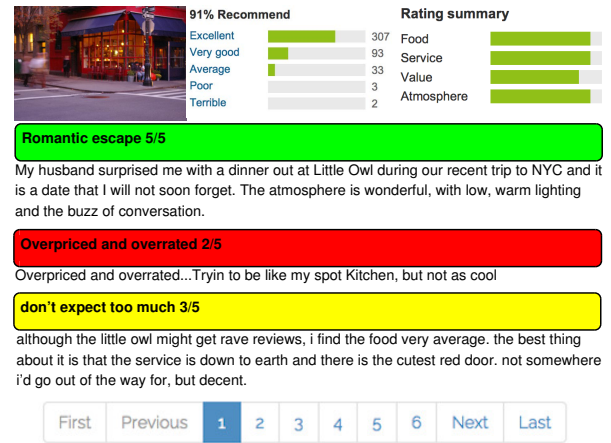


Figure 1: Restaurant page (a) with original TripAdvisor reviews and (b) with positive and negative summaries in place of reviews.

[3] G. Erkan and D. R. Radev. Lexrank: Graph-based lexical centrality as salience in text summarization. *J. Artif. Intell. Res.(JAIR)*, 22(1):457–479, 2004.

[4] Y. Gong and X. Liu. Generic text summarization using relevance measure and latent semantic analysis. In *Proc. of the 24th ACM SIGIR conference*, pages 19–25. ACM, 2001.

[5] L.-W. Ku, Y.-T. Liang, and H.-H. Chen. Opinion extraction, summarization and tracking in news and blog corpora. In *AAAI Spring Symposium*, volume 100107, 2006.

[6] C.-S. Lee, Z.-W. Jian, and L.-K. Huang. A fuzzy ontology and its application to news summarization. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 35(5):859–880, 2005.

[7] H. P. Luhn. The automatic creation of literature abstracts. *IBM Journal of research and development*, 2(2):159–165, 1958.

[8] R. Mihalcea and P. Tarau. Textrank: Bringing order into texts. In *Proc. of EMNLP*, volume 4, page 275. Barcelona, Spain, 2004.