Importance of Online Product Reviews from a Consumer's Perspective

Georg Lackermair[1,2], Daniel Kailer[1,2,*], Kenan Kanmaz[1]

[1]Munich University of Applied Sciences, Germany

[2]TU Dresden, Germany

*Corresponding Author: dkailer@hm.edu

Abstract   Product reviews and ratings are popular tools to support buying decisions of consumers. These tools are also valuable for online retailers, who use rating systems in order to build trust and reputation in the online market. Many online shops offer quantitative ratings, textual reviews or a combination of both. This paper examines the acceptance and usage of ratings and reviews in the context of e-commerce transactions. A survey among 104 German online shoppers was conducted to examine how consumer reviews and ratings are used to support buying decisions. The survey shows that reviews and ratings are an important source of information for consumers. However, qualitative feedback from the survey indicates that the perceived helpfulness of rating systems varies. Especially the comparison of user reviews is a very time consuming process for the customer, because of the unstructured nature of textual user reviews. In this paper we summarize similar problems and show corresponding examples to them. This will give new insight for future research in the area of user ratings and reviews.

Keywords   Product review, Rating, Empirical study, Comparability, Acceptance

aggregated to feedback profiles and published [11]. Such bi-directional communication models are described by the term Web 2.0 as defined by O'Reilly [5,6]. Popular examples of shopping websites containing user generated reviews and ratings are Amazon and eBay.

Due to the ubiquitous nature of product reviews, an interesting question is how customers actually use this source of information for their buying decision. To answer this question, a survey was conducted among 104 German online shoppers. The study was carried out in 2011 as part of a student's final project. The leading question of this work was: ,,How are user generated product reviews used for online transactions and what is their perceived importance for consumers?''.

Before answering this question, a short literature review in Section 2 will outline some aspects of previous research in the area of product reviews. Afterwards the concept and results of this empirical study will be presented in Section 3 and 4. As the results of the survey will show, product reviews are an important source of information for customers. However, qualitative feedback on the survey indicates that the perceived helpfulness of rating systems varies and that especially the comparison of user reviews is difficult. Therefore we will discuss some problems of product reviews in Section 5.

1 Introduction

Nowadays, more and more e-commerce platforms offer product reviews or product ratings. In literature, the terms review and rating are often used interchangeably, but for our work it is important to distinguish these two terms. A product review is a textual review of a customer, who describes the characteristics (e.g. advantages and disadvantages) of a product. A product rating on the other hand represents the customer's opinion on a specified scale. A popular rating scheme in online shops is the star-rating, where more stars indicate better ratings.

Product reviews and ratings are generated by the user (i.e. the customer of an online shop) and published on the website of the retailer. Additionally, the ratings are

2 Literature Review

Textual product reviews are often analyzed in text mining. Popescu and Etzioni [12] studied the extraction of semantic orientation and product features from such textual product reviews. Constant et al. [10] on the other hand studied the use of expressive content in various languages based on Amazon product reviews. The effects of Word-Of-Mouth (WOM) in online markets were also studied by various works. Dellarocas [11] discussed promises and challenges of online WOM networks. Chevalier and Mayzlin [9] studied the effect on sales of WOM-processes based on online book reviews.

The distribution of rating scores is also the subject of interesting research. Chevalier and Mayzlin [9] cal-

Table 1. Questions of the survey and their related constructs and hypotheses

Construct	Hypothesis	Question item	Choices
Trust	1	Q1: How important are online product reviews for your buying decision?	unimportant, rather unimportant, neutral, rather important, important
Decision making	2 (a)	Q2: Which typical sources of information do you use before you buy a product online?	Open question
	2 (b)	Q3: How often do you read product reviews before you buy a product online?	never, rarely, sometimes, regularly, often
	2 (c)	Q4: Do you usually compare positive and negative reviews?	yes, no
	2 (d)	Q5a: Was a single negative review ever the reason for you not to buy a product?	yes, no
		Q5b: Was a single positive review ever the reason for you to buy a product?	yes, no
Contribution	3	Q6a: Have you ever written an online product review?	yes, no
		Q6b: If yes, what was the overall rating of your review?	positive, neutral, negative

culated the distribution of ratings in online book reviews from Amazon, Barnes and Noble. The resulting data shows an asymmetric bimodal distribution (J-shaped distribution), whereas the most reviews are positive. Hu et al. [4] also calculated a J-shaped distribution of Amazon product reviews in three other product categories. They explain that this distribution is driven by self-selection bias – purchasing bias and under-reporting bias. Li and Hitt [7] studied the effect of self-selection bias on purchase decisions based on a set of Amazon product reviews.

3   Methodology

To understand how consumers use product reviews before buying online, a survey was conducted among 104 German online shoppers. The study was carried out in 2011 as part of a student's final project. The leading question of this work was: ,,Which role do user generated product reviews play in online transactions and what is their perceived importance for consumers?". To answer this question we derived several hypotheses for our survey.

Hypothesis 1 *Online reviews increase trustworthiness: Consumers consider the sole availability of user generated reviews and ratings as an important criterion for their evaluation of product attributes and the trustworthiness of a vendor.*

Hypothesis 2 *Reviews are an important source of information for online transactions: Consumers embrace reviews and ratings in the process of selecting products and vendors. For the indication of importance we used the following aspects: (a) awareness: users are aware of reviews and ratings and use the information purposely in their selection process; (b) frequency: online shoppers do frequently use reviews and ratings as source of information; (c) comparison: importance is indicated if consumers read reviews carefully and compare them with the opinion of other users; (d) effect: reviews and ratings can be considered important, if they affect the product selection.*

Hypothesis 3 *Users are willing to contribute: According to O'Reilly's [6] definition of Web 2.0, user generated reviews and ratings can be classified as services, whose quality improves as the contributions of users rise.*

Thus, an important aspect for reviews and ratings is consumers' willingness to contribute.

The main research question for the survey was: ,,How are user generated product reviews used in online transactions and what is their perceived importance for consumers?". To gather data about consumers' habits regarding online shopping, an online questionnaire was developed. The questionnaire design is summarized in Table 1 and includes the relationship between the questions and their assigned constructs and hypotheses.

Question item Q1 was included to find out whether online product reviews affect the buying decision of consumers. If consumers find reviews important or very important, they are more likely to trust a vendor if product reviews are present. The question items from Q2 to Q5 are connected to Hypothesis 2 and relate to the actual use of reviews. Question item Q2 was designed as an open question to determine the typical sources of information that consumers use before buying online. Question item Q3 intends to find out how often consumers read product reviews before buying online. Question item Q4 determines whether participants compare product reviews with each other. The question items Q5a and Q5b try to capture the consequences of positive and negative reviews. Finally, the question items Q6a and Q6b determine whether the participants create online reviews and what ratings they use.

The questionnaire was completed by 104 participants, where 53.85% were female and 45.19% were male. The remaining participants (0.96%) were not willing to share their gender. The age of the participants is distributed among five groups:

- 15–18 years old (2.88%)
- 19–26 years old (65.38%)
- 27–37 years old (24.04%)
- 36–50 years old (0.96%)
- Not willing to share their age (6.74%)

The participants where invited via email and online social networks. The demographics of the sample reflect the fact that the participants where acquired among the investigator's colleagues, fellow students and friends.

4   Results

The results of the survey are summarized in Table 2. The answers given in the questionnaire are aggregated as

relative frequency. Online reviews are rated as important or very important by 74.04% of the participants, whereas 4.80% think that reviews are rather unimportant or not important. The open question about the sources of information users seek before making a purchase, was answered differently. The answers indicate that most of the participants seek for advice online. Common answers include ,,search engine" (28%), ,,consumer reviews" (12%), ,,online shops" (14%) or ,,other online sources" (35%). Other online sources include blogs for example.

85.57% of the participants stated that they read reviews often or very often before they purchase online. Of the participants who read reviews, 83.65% compare positive and negative reviews with each other. 11.54% stated that a single negative review was the reason for not buying and 18.27% stated that a single positive review was the reason for buying a product. Among the participants, 57.69% have at least once written an online product review. 76.92% of these reviews were positive, 10.77% were negative and 12.31% had a neutral rating.

Table 2. Answers of the survey with their relative frequency

Question	Answer	Relative frequency
Q1	very important	.3077
	important	.4327
	neutral	.2115
	rather unimportant	.0192
	not important	.0288
Q2	search engine	.2800
	retail	.1200
	consumers' reviews	.1200
	online shops	.1400
	portals	.0900
	family and friends	.1500
	other online sources	.3500
	reviews in media	.0700
	commercials	.0200
	no investigation at all	.0500
	not specified	.1400
Q3	very often	.3942
	often	.4615
	sometimes	.0865
	rarely	.0365
	never	.0096
Q4	yes	.8365
	no	.1635
Q5a	yes	.1154
	no	.8846
Q5b	yes	.1827
	no	.8173
Q6a	yes	.5769
	no	.4135
	not specified	.0096
Q6b	positive	.7692
	neutral	.1231
	negative	.1077

5 Discussion

The results in the previous section demonstrated that user reviews are very important for the decision making of customers. But the more interesting and surprising finding was that a lot of customers compare positive and negative reviews with each other. In this section

we will therefore demonstrate on an exemplary use case, that current rating systems are not very efficient for the comparison of user reviews.

In our opinion there are two aspects of user reviews that play an important role. The first aspect regards the helpfulness of a written review. Customers read a full user review and expect to find something like a pro/con-list of features or opinions about a given product to support their buying decision. Customers can often rate a review as helpful or not helpful. An example for this is shown in Figure 1, which shows an excerpt of a user review from amazon.com and the rated helpfulness of the review.

Figure 1. Rated helpfulness of an amazon.com review
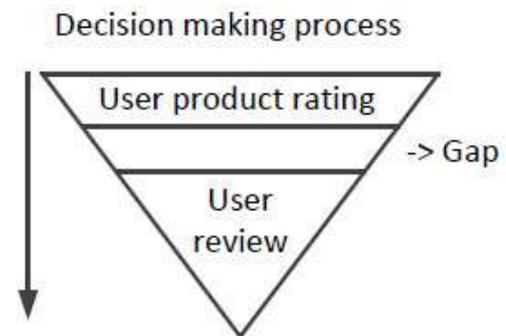
Figure 2. Decision making process, which displays the gap between user ratings and user reviews

A lot of research is done to predict the helpfulness of reviews, for example by Liu et al. [1], who built and evaluated algorithms for predicting the usefulness of IMDB movie reviews. Another study of Pan and Zhang [3] showed that the perceived helpfulness of a review positively correlates with the length of the review, which means that a long review text is usually more helpful than a short review text. But this also means that a potential customer has to read more text to find helpful information. This brings us to the second aspect, namely the efficiency of user reviews. Long reviews with a lot of information can be helpful, there is no doubt, but they are not always efficient for the customer. This is especially true for goal-oriented customers. Furthermore, it can also be very difficult to compare long product reviews.

We argue that the decision making process is divided into several phases. In an early phase, the customer wants to narrow down the products to get an overview of the potential products that match his requirements. We further argue that reading a lot of user reviews for several products is not very efficient in this early phase. This opinion was also confirmed by the qualitative feedback of the participants of our survey. Customers need compact and concise information about the products. In a later phase, when there are only a few potential pro-

Figure 3. Selected product reviews from amazon.com that contain the search term ,,running"

ducts left, the full user reviews can be used to get more details about the concrete products.

For the customer, a first indicator in the decision making process is often the user rating of a product, which is frequently expressed as a notation of stars, where more stars indicate better ratings. If a product has for example 100 user ratings with an average of 1 out of 5 stars, it is not very likely that the customer would even click on the product. But once customers enter the product page, they are dependent on reading user reviews to find out if the product matches their requirements. Therefore we argue that there is a gap in current rating systems between the user product rating and the user review, which is shown in Figure 2.

As we will see later, several approaches exist that are trying to bridge this gap. But at first we will look at an example to illustrate the problem of this gap. For a demonstration, we will look at a use case where a customer wants to buy an MP3-Player that is suitable for sports, especially for running purposes. Let us assume that the customer has found some MP3-Players with good ratings. The only way to find out if these products fit his purpose is to read the according reviews and see if any review mentions the running capabilities of the MP3-Player. Some sites may offer the possibility to search the reviews for a specific text. This is shown in Figure 3, where product reviews from amazon.com were searched for the word ,,running". But this search mechanism only shows where the word ,,running" appears. It is not very helpful for the user, because the results don't explicitly show whether the product is recommended for running or not.

Some sites try to overcome this issue by offering a more detailed product rating, compared to a single rating value as used by amazon.com (see Figure 3). The online shopping portal Ciao[1] for example uses multiple ratings as shown in Figure 4. Besides the main rating there are specific rating items that are common for a specific product group. In this case the detailed ratings are specific for an MP3-Player. Although this approach is more useful, it is still very limited, because the rating items are fixed (i.e. defined by the online store). This isn't helpful for the use case above, where the customer is looking for a suitable MP3-Player for running.

In the academic world, Vig et al. [2] provided a possible solution to above problem, called tag expression. Tag expression combines tags and ratings and allows users to apply affect to tags. For every tag, users can state if they like, dislike or are neutral about it. These tag expressions were tested with MovieLens[2], a movie recommendation website. The drawback of this approach is that only negative, neutral and positive ratings are possible. A more detailed rating, like the previously mentioned star-rating, needs also be tested with this approach. Furthermore the study only concentrated on a specific class of items (in their case movies). Therefore it needs to be evaluated whether an online shop with typical consumer products (like amazon.com) could also profit from this approach. Another consideration must be the review comparison feature, which was stated as important by the participants of our study, but was not examined in the work of Vig et al.

Another related research area is text mining, where product features and user opinions are extracted from user reviews in order to support buying decisions. But opinion extraction is still a hard problem not least because of the intrinsic complexity of languages [8]. One problem with opinion extraction is the mapping from an extracted opinion to a concrete rating value. For example, is the extracted opinion ,,This MP3-Player is great for running." equal to a 4 or 5 star rating? This question can't be answered by only extracting the text.

The examples show that existing rating systems do not support the decision making process seamlessly and force the consumers to read a lot reviews upfront. New approaches are needed to enable an easier summarization and comparison of product reviews.

As our study shows, it is important for the majority (about 83%) of the surveyed participants to compare positive and negative reviews. Comparing long and unstructured user reviews is a very time consuming process and therefore also limiting the efficiency of the decision making process. Comparing a more detailed rating like the one in Figure 4 would be more helpful, but isn't perfect, as we discussed earlier in our MP3-Player use case.

---

[1]Ciao: http://www.ciao.de

[2]MovieLens: http://movielens.umn.edu/
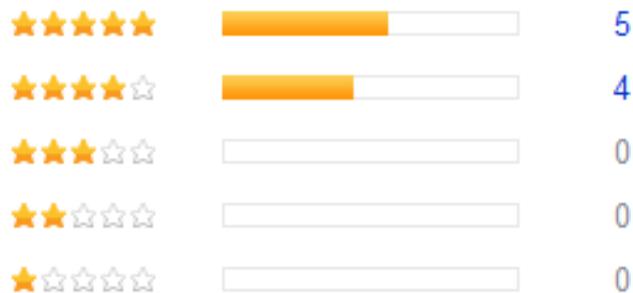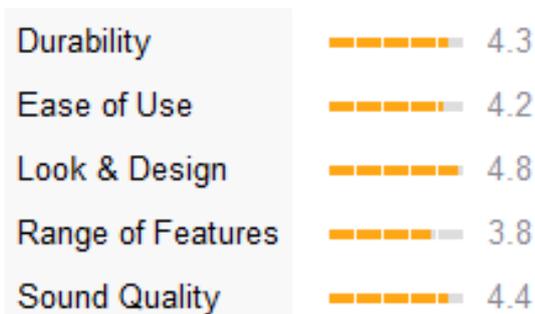
Figure 4. Ciao's detailed rating system, which displays detailed ratings for an MP3-Player

6   Conclusion and future work

The conducted survey from this article showed that product reviews are an important source of information for customers to support their buying decision. While product reviews can be very helpful for customers, the responses from the participants showed that current rating systems also have their weaknesses, especially when it comes to the task of comparing different product reviews with each other. We have demonstrated this problem with an exemplary use case, which shows that there is a gap between the product ratings and textual product reviews, which limits the efficiency of the decision making process of the customer.

As we have shown, many research projects try to predict the helpfulness of reviews. But in our opinion it is also important to research mechanisms to increase the efficiency of reviews to support the user in the decision making process. Such new approaches are also needed to allow an easier comparison of user reviews. The mentioned approach of tag expressions from Vig et al. [2] is interesting and will be considered in our future work. Currently we are working on a prototype that will allow a more detailed user rating, based on user generated tags. Additionally to existing research efforts, we will further investigate how these tag-based ratings can improve the comparison of reviews.

REFERENCES

[1] Y. Liu, X. Huang, A. An, X. Yu. Modeling and Predicting the Helpfulness of Online Reviews, Eighth IEEE International Conference on Data Mining, 442–452, 2008.

[2] J. Vig, M. Soukup, S. Sen, J. Riedl. Tag Expression: Tagging with Feeling, Proceedings of the 23nd annual ACM symposium on User interface software and technology, 323–332, 2010.

[3] Y. Pan, J. Q. Zhang. Born Unequal: A Study of the Helpfulness of User-Generated Product Reviews, Vol.87, No.4, 598-?612. 2011.

[4] N. Hu, P. Pavlou, J. Zhang. Overcoming the J-Shaped Distribution of Product Reviews, Communications of the ACM, Vol.52, No.10, 144–147. 2009.

[5] T. O'Reilly, What is Web 2.0. Design Patterns and Business Models for the Next Generation of Software, Online available from http://oreilly.com/web2/archive/what-is-web-20.html

[6] T. O'Reilly, Web 2.0 Compact Definition: Trying Again, Online available from http://oreilly.com/web2/archive/what-is-web-20.http://radar.oreilly.com/2006/12/web-20-compact-definition-tryi.html

[7] X. Li, L. M. Hitt. Self Selection and Information Role of Online Product Reviews, Information Systems Research, Vol.19, No.4, 456–474. 2008.

[8] C. Scaffidi, K. Bierhoff, E. Chang, M. Felker, H. Ng, C. Jin. Red Opal: Product-Feature Scoring from Reviews, Proceedings of the 8th ACM conference on Electronic commerce, 182–191, 2007.

[9] J. A. Chevalier, D. Mayzlin. The Effect of Word of Mouth on Sales: Online Book Reviews, Communications of the Journal of Marketing Research, Vol.43, No.3, 345–354. 2006.

[10] N. Constant, C. Davis, C. Potts, F. Schwarz. The pragmatics of expressive content: Evidence from large corpora, Communications of the Journal of Marketing Research, Vol.33, No.1-2, 5–21. 2009.

[11] Ch. Dellarocas. The Digitization of Word-Of-Mouth: Promise and Challenges of Online Reputation Mechanisms, Management Science, Vol.49, No.10, 1407–1424. 2003.

[12] A.-M. Popescu, O. Etzioni. Extracting Product Features and Opinion from Reviews, Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing , 339–346, 2005.