The Recipe for the Perfect Review?

An Investigation into the Determinants of Review Helpfulness

We analyze determinants of review helpfulness in online retailing based on Wang and Strong's (1996) data quality framework. Helpful reviews consist of 9 % of adjectives, display high product feature entropy, and present opinions that differ from previous reviews for the product in question. Other helpfulness determinants depend on whether experiential or utilitarian products are reviewed. Our research points e-shop providers towards two major improvements in their review systems.

DOI 10.1007/s12599-013-0259-3

The Authors

Prof. Dr. Michael Scholz
Juniorprofessur für E-Commerce
University Passau
Innstr. 43
94032 Passau
Germany
Michael.Scholz@Uni-Passau.De

Dr. Verena Dorner (✉)
Chair of Business Computing II
University Passau
Innstr. 43
94032 Passau
Germany
Verena.Dorner@Uni-Passau.De

Received: 2012-07-01
Accepted: 2013-02-07
Accepted after two revisions by Prof. Dr. Spann.
Published online: 2013-04-25

This article is also available in German in print and via http://www.wirtschaftsinformatik.de: Scholz M, Dorner V (2013) Das Rezept für die perfekte Rezension? Einflussfaktoren auf die Nützlichkeit von Online-Kundenrezensionen. WIRTSCHAFTSINFORMATIK. doi: 10.1007/s11576-013-0358-2.

Electronic Supplementary Material
The online version of this article (doi: 10.1007/s12599-013-0259-3) contains supplementary material, which is available to authorized users.

1 Introduction

Consumers' increasing propensity to shop online for all kinds of products – utilitarian and experiential – has made online retailing both a very attractive and a very competitive business. Online retailers are trying to attract and retain customers by offering innovative website features which improve the online "shopping experience". One particularly successful feature is the online review system, which provides consumers with the opportunity of exchanging their opinions on the products they bought. Consumer-generated product reviews decrease information search and product evaluation costs and reduce the incentive for consumers to leave the retailer's website in their search for product information. All major retailing websites now feature extensive review sections (e.g., Amazon.com, Staples.com). They have become so important that websites like Epinions.com have been able to make a successful business out of building and selling large "opinion bases" of consumer reviews.

But online review systems are threatened by their own success. Instead of alleviating information overload, they have become another source of it. Take the *Kindle Fire*, a tablet computer produced by Amazon. Since its release date on 15th November 2011, it has been reviewed over 18,000 times on Amazon.com. Even mundane items like Multipurpose Paper Reams impress with over 5,000 reviews (Staples.com). Most online retailers quickly realized that their customers need and want control beyond date-based sorting over the information displayed to them (Weathers et al. 2007) and introduced mechanisms to help consumers find good reviews faster. User-generated "helpfulness" rankings are perhaps the most popular control mechanism. Having read a review, consumers can vote on whether they thought it helped them in the purchasing process. Consumers can choose to have all reviews displayed in order of their helpfulness ratings. Perhaps not surprisingly, reviews with high helpfulness scores strongly influence product sales (Chen et al. 2008).

Unfortunately, helpfulness-based rankings suffer from two major flaws. First, many reviews are never voted on at all, which makes it impossible to rank them. Unranked reviews diminish the benefits consumers derive from using the review system: product information search and evaluation costs rise. Second, it is unclear which factors actually determine a review's helpfulness. In particular, current review systems do not provide mechanisms for estimating how valuable a new review is in relation to the set of currently available reviews. This makes it harder for reviewers to learn how to write good (i.e., helpful) reviews. Amazon.com, for instance, merely recommends that reviewers "be detailed and specific" and "aim for between 75 and 300 words" (the minimum length is 20 words). It is unclear on which evidence the Amazon.com advice is based. In fact, several empirical studies found a positive linear relationship between review length and helpfulness (Otterbacher 2008; Mudambi and Schuff 2010), suggesting rather "the longer the review the better".

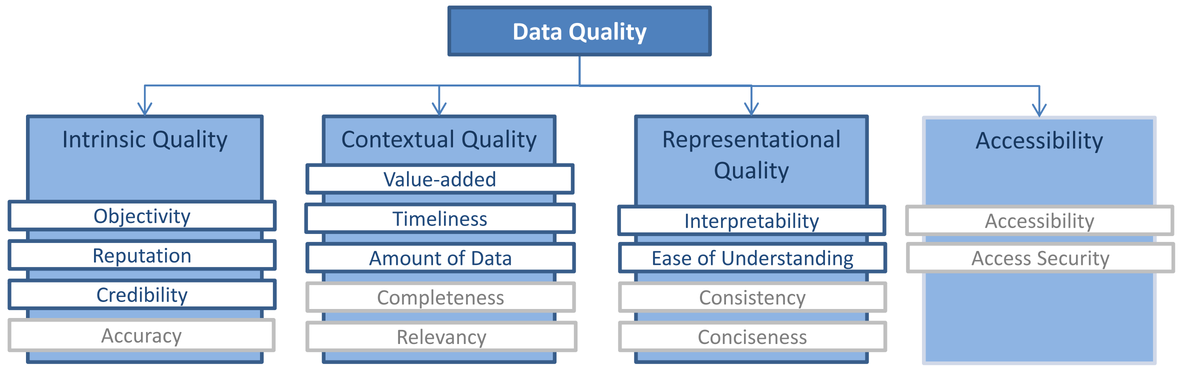

Fig. 1 Original data quality framework (Wang and Strong 1996)

We conducted a comprehensive survey of research on the determinants of review helpfulness in Amazon-like retailing review systems.[1] Unfortunately, the results reported in prior studies are contradictory. Our contribution to research on online consumer reviews is twofold. First, we build a basis for reconciling conflicting prior findings by anchoring them in Wang and Strong's (1996) well-known data quality framework. Second, we develop a model for explaining review helpfulness and empirically test it on three utilitarian and three experiential product categories from Amazon.com.

Review systems could be much improved if we knew which factors determine review helpfulness. Online retailers could introduce automatic review scoring systems to reduce consumer search costs. They could also provide better reviewer support to improve average review quality. Consumers would have less incentive to leave the retailer's website to search for information elsewhere and would find suitable products more easily, which would increase sales and customer satisfaction (Chen et al. 2008).

The remainder of this paper is organized as follows. In the next section, we review related work and show how Wang and Strong's (1996) data quality framework can be applied to structure prior research. Section 3 describes our research model for investigating the determinants of review helpfulness. Section 4 presents the empirical evaluation of our research model. Implications for research and practice as well as limitations follow in Sect. 5.

2 Literature Review and Theoretical Foundations

2.1 Framework of Review Helpfulness Determinants

Research on traditional (offline) word-of-mouth (WOM) has produced a number of explanations why consumers might find a particular piece of information helpful. Many of these have been transferred to and tested in online (electronic WOM, eWOM) contexts (Dellarocas 2003). Among the features which have been found to affect readers' perceptions of online product reviews are, for instance, writing style, number of discussed product attributes, and author identity disclosure (Ghose and Ipeirotis 2011). But prior research has produced mixed results regarding the influence of these features. A structured approach which would help explain and reconcile these differences is missing so far. We applied Wang and Strong's (1996) data quality model to discover the main categories of helpfulness determinants. High-quality or "helpful" reviews are those which help online consumers purchase "the right" product. This understanding are consistent with Wang and Strong's (1996) definition of high-quality data as "data that are fit for use by data consumers", with "fit" being determined by the (data) needs of the "data consumer" for solving a specific task.

All previously investigated review "fitness" or helpfulness determinants fall within three out of the four original dimensions, "intrinsic quality", "contextual quality", and "representational quality" (Wang and Strong 1996).[2]

Online-Appendices A and B present previous research structured according to our adapted model (**Fig. 1**).

2.2 Existing Literature on Review Helpfulness Determinants

Objectivity, while desirable in advertising where consumers are less inclined to believe subjective than objective information (Ford et al. 1990), can have quite the opposite effect in the eWOM context. Readers of (largely anonymous) reviews try to infer from review content to which extent their preferences overlap with the reviewer's preferences. Product-related statements and opinions may be influenced either by the reviewer's personal preferences or by "environmental" causes, in our case product features (Folkes 1988). Subjective reviews – discussing personal experiences with the product – permit easy inferences as to the causes (environmental or personal) for evaluative statements, and to how likely it is that the reader's and the reviewer's product assessments will coincide. Results from prior research are mixed. Zhang and Varadarajan's (2006) and Liu et al. (2007) results suggest that subjectivity performs poorly in separating helpful from unhelpful reviews. Hao et al. (2009) and Ghose and Ipeirotis (2011), however, found that subjectivity significantly influences helpfulness. Chen and Tseng (2011) could distinguish best between helpful and unhelpful reviews with models including this feature.

Another signal that a review might contain valuable, true information is the author's reputation. If consumers are not

[1] Amazon-like review systems require reviewers to give an overall star rating and a free-format review (minimum length 20 words), but not to rate products on mandatory standardized criteria (e.g., TripAdvisor.com).

[2] Two studies on review helpfulness (Jin and Liu 2010; Chen and Tseng 2011) used parts of this framework to derive review quality measures.

motivated or cognitively able to process elaborate messages, superficial peripheral cues can improve message persuasion (Cacioppo and Petty 1984). One such cue is authority. Reviewers can enhance their authoritative status by posing as (self-described or third-party verified) experts for a certain product category; or they can establish a reputation for writing particularly helpful reviews. Apart from reducing the cognitive costs of processing reviews, such signals also help consumers infer whether a reviewer's evaluation is based on non-veridical knowledge (Eagly and Chaiken 1984). Results by Liu et al. (2008) and Chen and Tseng (2011) indicate that reputation (in combination with other features) is a very good classifier.

Consumers draw similar inferences about source credibility, which influences review persuasiveness and adoption, especially when consumers evaluate experiential goods (Jain and Posavac 2001). Credibility in our context refers to the extent to which a review can be trusted to contain true information. This may not be the case if the reviewer is reluctant to relate her true experiences or beliefs (Eagly and Chaiken 1984). Consumers may try to infer whether a review is likely to contain truthful information by checking the reviewer's identity, e.g., to discover ties to the reviewed product's company. Forman et al. (2008) found that reviewers who disclose their name tend to receive a higher number of helpfulness votes on their reviews, in particular when their overall product rating is moderately equivocal. Ghose and Ipeirotis's (2011) results also indicate that identity disclosure positively affects perceived helpfulness.

Value-added indicates whether a new review provides additional or different opinions compared to the opinions already available from the review corpus. This can serve to adjust early reviewer bias (if present) over time (Li and Hitt 2008) and increase consumer trust by providing a broader scope of opinions (Schlosser 2011). If extreme ratings dominate review sections for certain products, consumers may think it unlikely that any reviews contain balanced arguments. Consumers who are interested in "both sides of the coin" will feel the need to refer to a variety of reviews to collect diverse opinions. If prior reviews are all positive (negative), consumers are likely to perceive a new review with a negative (positive) opinion as more helpful than another positive (negative) review. Results by Danescu-Niculescu-Mizil et al. (2009), Jin and Liu (2010), and Otterbacher (2008) support the supposition that divergent reviews add value to the review corpus.

The amount of data contained in a review has a positive impact on perceived helpfulness (Zhang and Tran 2011). Mudambi and Schuff (2010) account for this finding with the explanation that decision confidence is higher when more reasons are made available to the decision-maker (Tversky and Kahneman 1974), and that arguments are perceived as more persuasive when the quantity of available information is greater (Schwenk 1986). Regardless of product type, apparently, a higher amount of data in the review increases decision confidence and review persuasiveness and has a positive impact on helpfulness (Korfiatis et al. 2008; Mudambi and Schuff 2010; Zhang et al. 2010; Wu et al. 2011).

Timeliness describes how early a review was posted. This may influence perceived helpfulness in one of two ways. Earlier reviews may attract more attention and more positive helpfulness votes. Later reviews, on the other hand, may be viewed as more up-to-date and, for this reason, as more helpful (Hao et al. 2009). Results on the impact of timeliness are mixed: Otterbacher (2008) and Chen and Tseng (2011) found that it does not significantly improve classification. Wang et al. (2011) included a time-decay factor in computing helpfulness-based review rankings and found that their metric outperforms purely helpfulness score-based rankings. Pan and Zhang (2011) found that a review's age has a significant positive effect on review helpfulness.

Ease of understanding positively influences the likelihood that a consumer will adopt a recommendation (Eagly 1974) and that she will perceive a review as helpful (Ghose and Ipeirotis 2011; Chen and Tseng 2011). Lower levels of ease of understanding indicate that higher cognitive effort is necessary to process a text, which makes it (on average) less likely that a random reader will understand it. Easily comprehensible reviews, which require less cognitive effort in reading, are more likely to be perceived as helpful (Korfiatis et al. 2008; Ghose and Ipeirotis 2011; Wu et al. 2011).

Interpretability refers to the degree to which the opinions voiced in a review are easily identified as being positive or negative. If they are easy to interpret (classify) as positive or negative, the review is likely to be perceived as helpful (Schlosser 2011). Reviews are particularly helpful if they contain highly diagnostic information which helps consumers assign a product to precisely one cognitive category (for instance, "low quality"). Review diagnosticity is expressed by the star rating. Extreme ratings permit the consumer to classify the reviewed product almost instantly as "good" or "bad". Reviews with moderate overall ratings are more likely to contain ambiguous information. Evaluating such reviews demands a greater cognitive effort, which leads to lower perceived helpfulness (Forman et al. 2008; Ghose and Ipeirotis 2011). Contrary to these findings, Pan and Zhang (2011) discovered positive linear relationships between rating and helpfulness for both experiential and utilitarian products. Reviews on experiential products were generally considered less helpful, and negative reviews on experiential products were significantly less likely to be helpful than negative reviews on utilitarian goods. Mudambi and Schuff (2010) reported yet another result: they found that product type moderates the relationship between rating extremity and helpfulness. Experiential goods exhibited an "inverted U"-shaped relationship and utilitarian goods a "U"-shaped relationship. Mudambi and Schuff (2010) argue that consumers of experiential goods prefer moderate to extreme reviews because they are interested in obtaining balanced information. If the assumption that consumers expect moderate reviews to hold more balanced views is true, the overall rating acts as a peripheral cue for review balance. Since it is harder for consumers of experiential goods to infer from review content whether personal or product-related causes shape the reviewer's opinions, they find balanced reviews more helpful. Consumers of utilitarian goods, on the other hand, may prefer extreme reviews because they can easily discern personal and product-related causes for reviewer (dis-)satisfaction.

Structuring all prior research along the lines of Wang and Strong's (1996) framework thus provides a comprehensive view of the factors (potentially) determining review helpfulness (**Fig. 1**), which we will use in Sect. 3 to build our research model. In the next subsection, we discuss how we adapted Wang and Strong's (1996) generic framework to fit it to our context. Online-Appendix C summarizes the adapted definitions for all relevant data quality constructs.

2.3 Adapted Framework of Review Helpfulness Determinants

The dimension "intrinsic quality" contains all single-review characteristics related to review content (objectivity, accuracy) and author (reputation, credibility). Accuracy, albeit intuitively important, is difficult to measure. Divergent opinions on product features may not indicate errors in judgment but be due to different consumer preferences (Li and Hitt 2008), expectations (Bone 1995), or even isolated manufacturing errors. The construct accuracy must be dropped from this dimension for lack of reliable measurement instruments.

"Intrinsic quality" assumes that quality depends solely on review-immanent features. In contrast, the dimension "contextual quality" sets each review in relation to its context, which is described by reader requirements (relevancy) and review corpus (value-added, timeliness, appropriate amount of data, completeness). Wang and Strong's original dimensions only take data consumers' task-related requirements into account. Since the specific information-seeking task is unknown in our context (e.g., birthday present or manufacturing problem), the task-centric perspective is a poor fit to our research setting. We therefore propose that relevancy be omitted. Many previous studies consider reviews which discuss all features as the gold standard for helpfulness. We believe that this view systematically underestimates the value of reviews which evaluate a product only partially. If no other review has commented on a particular feature before, a review focusing on this feature may be more helpful than another comprehensive review which does not address this feature. Completeness may not even be desirable for consumers who search for information on particular features only. In our opinion, it is a secondary concern in review systems that quickly attract large numbers of reviews and is therefore omitted from our analysis.

"Representational quality" contains all aspects related to review formatting (conciseness, consistency) and writing (interpretability, ease of understanding). Since all major retailers only offer one format type which is determined by the retailer, conciseness and consistency are largely irrelevant for all research conducted on a single website.

As a final adaptation, we propose to drop the dimension accessibility.[3] Data management-related issues like data security and confidentiality cannot be influenced by the reviewer, but are platform-dependent. We therefore omit them from further discussion.

Prior research has not addressed three problems so far. First, almost all previously used prediction models do not accommodate all (potentially) important review features (see Online-Appendix A), reducing their model's explanatory power. Chen and Tseng's (2011) study is the only exception; however, their main contribution lies in evaluating the performance of different support vector machines and feature-based classifiers for query-dependent review retrieval. The results are only partially suitable for explaining which features affect review helpfulness to which degree.

Second, most studies do not differentiate between product type. The few who do, in particular Mudambi and Schuff (2010) and Pan and Zhang (2011), provide evidence for moderating effects of product type on a number of relationships between review features and review helpfulness.

Third, many studies employ data collection strategies that could produce a biased data base. One of the most common strategies is collecting only the reviews for a certain number of bestselling products (e.g., Mudambi and Schuff 2010; Chen and Tseng 2011; Pan and Zhang 2011). Consumers, however, react differently to hit and niche products (Berger et al. 2010; Dellarocas et al. 2010): the data may be afflicted with a "bestseller bias". Another potentially bias-inducing strategy is choosing only the most recent reviews (e.g., Chen and Tseng 2011): age has been shown to influence helpfulness perceptions (Otterbacher 2008; see Sect. 2.2).

In the next section, we present our research model. It addresses the abovementioned issues by (i) including the salient determinants in all categories and (ii) accounting for product type. In Sect. 4, we also address the third issue by employing a parsimonious data collection strategy.

3 Research Model

In the "intrinsic quality" dimension, review objectivity, author reputation, and author credibility are the main determinants of review helpfulness (e.g., Park et al. 2007; Hao et al. 2009; Ghose and Ipeirotis 2011; Chen and Tseng 2011).

We suggest that the mixed results for objectivity are, at least partly, due to the lack of differentiation between product types. Subjective reviews, relating personal product experiences, are likely to be considered as particularly valuable for experiential goods. Consumers are almost entirely dependent on information about the reviewer's personal preferences and product usage to successfully infer whether they will be (dis-)satisfied with the product. Most attributes of utilitarian products, on the other hand, can be described objectively. We suggest that objective reviews on a utilitarian product's performance are more helpful because they permit consumers to assess the truth of the manufacturer's claims.

H1a: *Objective reviews are more helpful for utilitarian products than for experiential products, whereas subjective reviews are more helpful for experiential than for utilitarian products.*

A reviewer's reputation depends on the number of consumers who have read her previous reviews and found them helpful. Reviewers who were able in the past to write helpful reviews in a product category ("good reviewers") are likely to be able to do so again in the future. In addition, consumers sometimes acquire preferences for certain authors whose opinions or style they like (Burton and Khammash 2010) and give their reviews more (positive) attention. The probability of a reviewer obtaining positive votes on future reviews increases with the quality of her previously published reviews (i.e., her reputation in the review system).

H1b: *Perceived review helpfulness increases with the average helpfulness of the author's previous reviews in the same product category.*

According to Forman et al. (2008), consumers use "disclosure" as a criterion for choosing which reviews to read, with "disclosure" acting as a cue for message credibility. We suggest that a review whose author's real identity is known will be perceived as more helpful by consumers.

[3] The original framework was developed in a business/enterprise database management context.

H1c: *Product reviews from authors that have disclosed their identity are more helpful to consumers.*

The "contextual quality" features that influence review quality strongly are value-added, timeliness, and appropriate amount of data (e.g., Danescu-Niculescu-Mizil et al. 2009; Pan and Zhang 2011; Jin and Liu 2010). A review can add value by expressing a dissenting view from the majority of existing reviews: consumers are offered a broader spectrum of opinions (Danescu-Niculescu-Mizil et al. 2009).

H2a: *Reviews that deviate strongly from previous reviews are perceived as more helpful than reviews that conform with previous reviews.*

Consumers evaluate the quality of products by evaluating product attributes (Netzer and Srinivasan 2011; Scholz et al. 2010). Product reviews that discuss many attributes that a particular consumer is interested in or that are among the first to discuss an attribute are therefore likely to be helpful for that consumer.

H2b: *Perceived review helpfulness increases with the amount of information.*

We suggest that earlier reviews will be considered more helpful because they attract more attention and, hence, more positive helpfulness votes (Pan and Zhang 2011).

H2c: *The earlier a review is posted, the more helpful it is to consumers.*

The third dimension "representational quality" includes the helpfulness determinants interpretability and ease of understanding (e.g., Korfiatis et al. 2008; Ghose and Ipeirotis 2011; Wu et al. 2011). Reviews which require less cognitive effort in reading are likely to be better understood by consumers and to be perceived as more helpful than reviews that are difficult to read (Ghose and Ipeirotis 2011).

Past research in the field of business and information systems engineering (BISE) developed and used tests for identifying the readability based on the (grade) level of education that is needed to comprehend a piece of text (Klare 2000; Ghose and Ipeirotis 2011). The lower the level of education required to understand a review, the larger the number of consumers who are able to comprehend it.

H3a: *Perceived review helpfulness increases with higher levels of readability.*

Readability indices are formulae for estimating the readability of a text based on the relative frequencies of characters, words, sentences, or syllables. Two short reviews with simple words have the same readability index value even if one of them is full of spelling errors, which impair readability. Supplementing readability indices with the number of spelling errors permits better evaluation of review understandability (Klare 2000; Ghose and Ipeirotis 2011).

H3b: *Perceived review helpfulness increases with lower numbers of spelling errors.*

Interpretability refers to the ease with which an author's opinion can be extracted from a review. Both highly diagnostic reviews (Mudambi and Schuff 2010) and highly "accessible" reviews (Feldman and Lynch 1988) are easily interpretable and therefore likely to be helpful. "Accessible" reviews are easily recalled from memory because they discuss the product particularly engagingly or provokingly (Herr et al. 1991). Writing vivid product descriptions or provocative statements requires using many adjectives or adverbs. Indeed, sentiment analysis models perform much better in identifying the polarity of a text and extracting its author's opinions when they include adjectives and adverbs (Benamara et al. 2007; Xia et al. 2011). Of course, increasing the proportion ad infinitum is unlikely to improve review quality. Using too many adjectives and adverbs in relation to other words (product features, verbs, etc.) renders a review less comprehensible and less diagnostic. We propose that a moderately high proportion of adjectives and adverbs improves both review diagnosticity and accessibility and therefore review helpfulness.

H3c: *Product reviews with extremely high or extremely low adjective (adverb) frequency are less helpful than reviews with moderately high adjective (adverb) frequency.*

If the assumption that consumers expect moderate reviews to hold more balanced views is true, consumers may use the overall rating as a peripheral cue for review balance. It is more difficult for consumers of experiential goods to infer from review content whether personal or product-related causes shape the reviewer's opinions. These consumers will consider balanced reviews more helpful. Consumers of utilitarian goods, on the other hand, can easily discern personal and product-related causes for reviewer (dis-)satisfaction and will prefer extreme reviews (Mudambi and Schuff 2010).

H3d: *Product reviews with extreme ratings are less helpful than reviews with moderate ratings if the reviewed product is an experiential good, whereas reviews with extreme ratings are more helpful than reviews with moderate ratings if the reviewed product is a utilitarian good.*

The complete research model is summarized in **Fig. 2**. We describe the empirical evaluation in the following Sect. 4.

4 Empirical Evaluation

4.1 Data Collection

We collected data for 1006 digital compact cameras, 1402 smartphones, 569 notebooks, 133 daypacks, 1507 board games, and 182 eaux de toilette available from Amazon.com in March 2012. We decided to collect data from these six categories because (i) most products in these categories have been reviewed, (ii) they comprise both utilitarian and experiential products and (iii) the products cover a wide price range. To avoid introducing selection biases (Sect. 2.3), we collected all reviews for all products in the chosen product categories. To the best of our knowledge, this data collection is the most diverse and extensive review corpus so far.

We excluded products without reviews and products with identical reviews (e.g., technically identical cameras of different color) and reviews with less than ten votes from our analysis in order to ensure robust approximations of review helpfulness. Overall, we analyzed 27,104 reviews across the six product categories. For each review, we collected helpfulness as number of positive votes/number of total votes, overall star rating, review date, review text and reviewer name, and whether the reviewer's real name is given. Rating deviation is computed as the difference between a review's star rating and the average rating of all prior reviews for the same product. Positive deviations indicate that later buyers are more satisfied with the product than early buyers. Reviewer reputation is measured by the average helpfulness of one reviewer's prior reviews in the same product category. Objectivity is computed as 1 minus

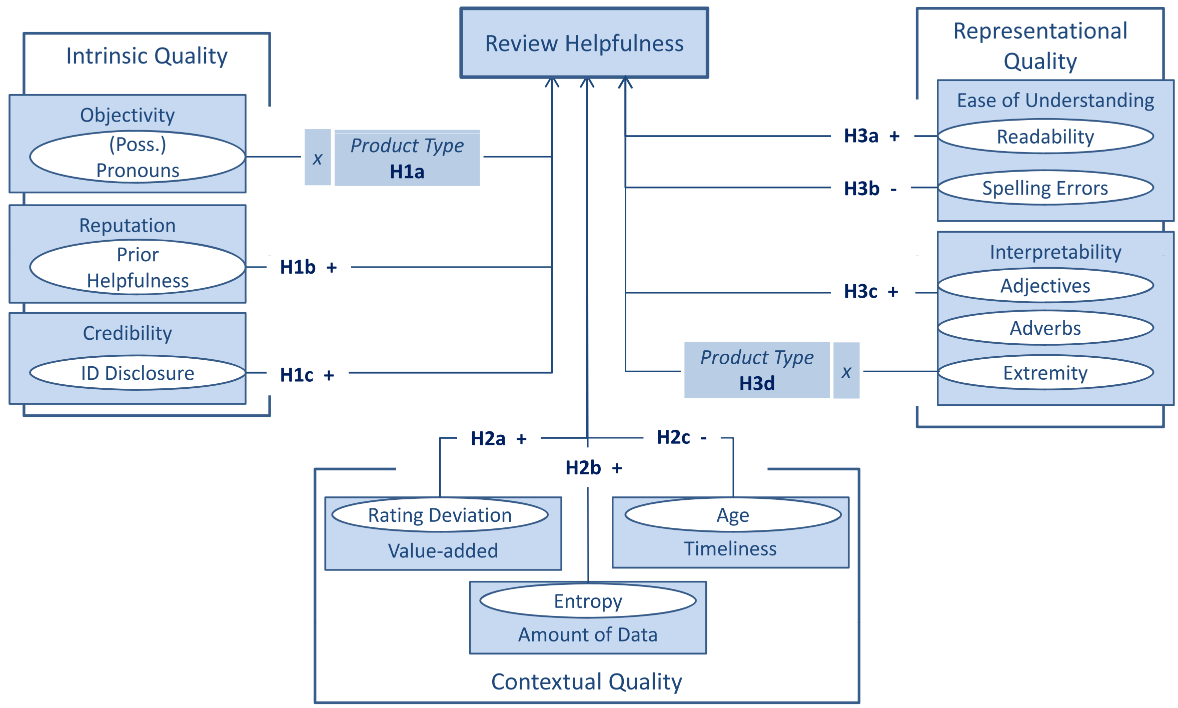

Fig. 2 Research Model

the proportion of personal and possessive pronouns.

We identified sentences, words, syllables, and part-of-speech. For estimating review readability, we chose SMOG, which is a very robust measure (Lahiri et al. 2011). To compute SMOG, we counted the number of sentences and the number of polysyllables (words of 3 or more syllables). The SMOG index was then computed as

$$\begin{aligned}\mathrm{SMOG} = 1.043 \\ &\times \Big(\text{number of polysyllables} \\ &\times \frac{30}{\text{number of sentences}}\Big)^{\frac{1}{2}} \\ &+ 3.1291\end{aligned}$$

For each of the six product categories, we defined a set of product features.[4] To find out which features are discussed in a particular review, we compared all nouns and proper nouns with this set. We used Shannon's information entropy concept (Shannon 1948) to measure the amount of information contained in a review (Otterbacher 2008; Zhang and Tran 2011). A review's entropy is computed as

$$-\sum_{\forall\, features} p(feature) \log_2\big(p(feature)\big)$$

where $p(feature)$ is the probability of a product feature being discussed in a review.

We performed a spell-check, based on wiktionary.org, on all words except the proper nouns. **Table 1** presents descriptive statistics for all variables.

The descriptive statistics for experiential and utilitarian goods exhibit several interesting patterns. On average, reviews for experiential products are considered more helpful than reviews for utilitarian products. One reason for this difference might be the fact that board games and eaux de toilette posses very few predefined features. This means that the amount of easily accessible information is very small and pre-purchase uncertainty is very high (Mudambi and Schuff 2010). In contrast, our utilitarian goods possess many well-defined features which can be evaluated largely without additional information (Nelson 1970). Reviews for utilitarian goods discuss more features than reviews for experiential goods, are of greater length, and display higher levels of entropy than experiential goods' reviews. But the discussion of such a large number of features, and possibly their interactions, is more difficult to comprehend. This is evidenced by these reviews' significantly worse readability scores (**Table 1**).

While there exists a positive linear relationship between number of words and number of features discussed ($p < 0.001$), we discovered an "inverted U"-shaped relationship between number of words and product rating. Moderate reviews (3 stars) tend to be "balanced", i.e., discuss both pros and cons, and are consequently longer than extreme reviews.

4.2 Analysis and Results

Our dependent variable is the probability of a review being helpful. Logistic regressions are appropriate for binomial variables if and only if the number of trials and successes of the dependent variable are known (Wright and London 2009, p. 94). In our case, as Mudambi and Schuff (2010) note, the number of trials, i.e., of consumers who have read a particular review, is unknown: a review's true helpfulness may differ from its revealed helpfulness. In such cases, Tobit regressions are more robust than logistic re-

[4]The product features were collected from the manufacturers' homepages.

Table 1 Descriptive statistics of our sample [mean (standard deviation)]

Measure	Camera	Smartphone	Notebook	Daypack	Board game	Eau de toilette
Type	Utilitarian	Utilitarian	Utilitarian	Experiential	Experiential	Experiential
# Products	1006	1402	569	133	1507	182
# Reviews	8424	13047	1701	64	2607	1261
Words per review	243.95 (251.21)	267.18 (297.53)	278.58 (338.07)	114.94 (119.51)	173.68 (208.07)	193.36 (227.69)
Helpfulness	0.70 (0.28)	0.62 (0.32)	0.68 (0.30)	0.84 (0.23)	0.76 (0.27)	0.72 (0.28)
Readability	9.03 (3.33)	8.64 (3.32)	8.76 (3.51)	7.76 (4.10)	7.74 (2.62)	8.63 (3.51)
Spelling errors (%)	13.76 (7.20)	13.59 (7.26)	14.22 (6.87)	10.07 (5.30)	10.92 (8.94)	14.14 (7.95)
Adjectives (%)	6.04 (3.11)	5.12 (2.90)	5.27 (2.76)	5.72 (3.61)	5.41 (3.03)	5.47 (3.20)
Adverbs (%)	6.44 (3.59)	6.60 (3.67)	6.73 (3.69)	7.35 (4.24)	5.54 (3.71)	6.80 (3.93)
Star rating	3.77 (1.43)	3.42 (1.59)	3.81 (1.44)	4.19 (1.25)	3.88 (1.38)	3.66 (1.50)
Entropy	2.32 (2.56)	1.83 (2.12)	3.19 (3.45)	0.51 (0.68)	0.74 (0.92)	0.19 (0.39)
Age in days	1145.84 (803.65)	979.50 (680.88)	664.18 (509.74)	679.23 (392.87)	1904.15 (831.94)	1692.25 (747.89)
Rating deviation	−0.04 (1.97)	−0.45 (1.98)	0.44 (2.44)	1.16 (2.69)	0.72 (2.54)	−0.26 (1.52)
Objectivity	0.91 (0.10)	0.86 (0.16)	0.86 (0.18)	0.93 (0.08)	0.93 (0.08)	0.91 (0.11)
Prior helpfulness	0.25 (0.37)	0.22 (0.34)	0.15 (0.30)	0.03 (0.15)	0.30 (0.39)	0.08 (0.23)
Reviewers with prior reviews in this category (%)	35.92	34.64	21.22	3.13	37.82	11.34
ID disclosure (%)	31.17	30.63	25.57	35.94	32.34	38.30
Predefined features	71	83	75	51	22	16

gressions – if most data do not take on the values of the dependent variable's interval limits. Since 90 % of our dataset (24,407 out of 27,104 observations) do not equal the limit values of 0 and 1, we may assume that the Tobit regression results are unbiased (Greene 2012, p. 895). Our regression model, including all interaction effects, is summarized by the following equation.

$$\begin{aligned}
\mathit{Helpfulness} &= \alpha + \beta_1 \mathit{Objectivity} + \beta_2 \mathit{Objectivity} \\
&\quad \times \mathit{Product\ Type} \\
&\quad + \beta_3 \mathit{Prior\ Helpfulness} \\
&\quad + \beta_4 \mathit{Identity\ Disclosure} \\
&\quad + \beta_5 \mathit{Rating\ Deviation} \\
&\quad + \beta_6 \mathit{Entropy} + \beta_7 \ln(\mathit{Age}) \\
&\quad + \beta_8 \mathit{SMOG} + \beta_9 \mathit{Spelling\ Errors} \\
&\quad + \beta_{10} \mathit{Adjectives} + \beta_{11} \mathit{Adjectives}^2 \\
&\quad + \beta_{12} \mathit{Adverbs} + \beta_{13} \mathit{Adverbs}^2 \\
&\quad + \beta_{14} \mathit{Rating} \\
&\quad + \beta_{15} \mathit{Rating}^2 + \beta_{16} \mathit{Rating} \\
&\quad \times \mathit{Product\ Type} + \beta_{17} \mathit{Rating}^2 \\
&\quad \times \mathit{Product\ Type} \\
&\quad + \beta_{18} \mathit{Product\ Type} + \varepsilon
\end{aligned}$$

Table 2 presents the results of the Tobit regression. Inter-variable correlations of at least 0.25 indicate absence of multicollinearity. A highly significant likelihood ratio test ($p < 0.001$) and a pseudo-R^2 value of 0.671 indicate an outstanding model fit.

We propose that a review's helpfulness is influenced by its representational, contextual and intrinsic quality.

Intrinsic review quality is operationalized with the number of (possessive) pronouns in relation to review length (objectivity), an author's prior reviews' helpfulness (reputation) and identity disclosure (credibility). Separate Tobit regressions for utilitarian and experiential goods, testing the hypothesized moderating effect of product type on the relationship between objectivity and helpfulness (H1a), revealed that consumers prefer objective reviews for utilitarian and subjective reviews for experiential goods. This could explain the mixed results reported by previous studies (e.g., Liu et al. 2007; Hao et al. 2009; see Online-Appendix B): most did not distinguish between product types or used only one type of product. We used the following equation in the additional regression analyses.

$$\begin{aligned}
\mathit{Helpfulness} &= \alpha + \beta_1 \mathit{Objectivity} \\
&\quad + \beta_2 \mathit{Prior\ Helpfulness} \\
&\quad + \beta_3 \mathit{Identity\ Disclosure} \\
&\quad + \beta_4 \mathit{Rating\ Deviation} + \beta_5 \mathit{Entropy} \\
&\quad + \beta_6 \ln(\mathit{Age}) + \beta_7 \mathit{SMOG} \\
&\quad + \beta_8 \mathit{Spelling\ Errors} + \beta_9 \mathit{Adjectives} \\
&\quad + \beta_{10} \mathit{Adjectives}^2 + \beta_{11} \mathit{Adverbs} \\
&\quad + \beta_{12} \mathit{Adverbs}^2 + \beta_{13} \mathit{Rating} \\
&\quad + \beta_{14} \mathit{Rating}^2 + \varepsilon
\end{aligned}$$

Reviewer reputation has a significant positive impact on helpfulness (H1b), whereas identity disclosure (H1c) does not. Consumers apparently do not use "identity disclosure" as a cue for review credibility.

Contextual quality factors (value added to the review corpus, amount of information relative to previously available information, and timeliness) have a significant influence on review helpfulness. Rating deviation (H2a) and entropy (H2b) both have a significant positive effect on review helpfulness. Consumers evidently appreciate being provided with diverse opinions. Support for H2b indicates that consumers value reviews that discuss many attributes as well as reviews that discuss attributes which have not been mentioned by other reviewers before. The age

Table 2 Regression results

Variable	Estimate	Standard error	Hypothesis	
Constant	25.246***	0.891		
Objectivity	−0.592***	0.047	*Objectivity*	H1a (supported)
Objectivity × product type	0.302***	0.048		
Prior helpfulness	0.081***	0.005	*Reputation*	H1b (supported)
Identity disclosure	−0.002	0.003	*Credibility*	H1c (not supported)
Rating deviation	0.014***	0.001	*Value-added*	H2a (supported)
Entropy	0.028***	0.001	*Amount of Data*	H2b (supported)
Log(age)	0.879***	0.032	*Timeliness*	H2c (supported)
SMOG	0.014***	0.001	*Readability*	H3a (not supported)
Spelling errors in %	−0.152***	0.023	*Spelling Errors*	H3b (supported)
Adjectives in %	1.613***	0.111	*Interpretability*	H3c (supported)
Adjectives in %^{2}	−9.085***	0.591		
Adverbs in %	0.344***	0.092		
Adverbs in %^{2}	−3.554***	0.383		
Rating	0.141***	0.018	*Extremity*	H3d (not supported)
Rating^2	−0.013***	0.003		
Rating × product type	−0.065***	0.019		
Rating^2 × product type	0.010***	0.003		
Product type	−0.302***	0.052		
Log-likelihood	−4425.478***			
Pseudo-R^2	0.673			

of a review (H2c) is positively correlated with helpfulness, indicating that more recent reviews are less helpful.

Finally, representational quality is operationalized with readability, number of spelling errors (ease of understanding), proportion of adjectives and adverbs to other words, and overall rating extremity (interpretability). Our results indicate that less easily readable reviews (higher SMOG values) are more helpful, contrary to H3a.[5] At a first glance, this finding appears counterintuitive. But a closer look at SMOG index suggests two interesting explanations. SMOG is computed based on the proportion of polysyllables to sentences. Hence, reviews with a high ratio of polysyllabic (i.e. difficult) words are perceived as more helpful. Many of the polysyllables used in the reviews are product features, adjectives and adverbs (e.g., resolution, megapixel, technically). This suggests, firstly, that consumers value attribute-based, comparatively complicated "technical" reviews. Second, consumers might gloss over words they do not understand (i.e. attributes whose functionality is unclear to them) and merely "add up" the pros and/or cons. Reviews that discuss many attributes would be particularly valuable to such consumers. This interpretation is supported by the fact that high review entropy (H2b) has a significant effect on review helpfulness and that SMOG and entropy are moderately but significantly correlated ($r = 0.17$). The number of spelling errors (H3b) has a significant negative effect on helpfulness.

The relationship between helpfulness and the number of adjectives or adverbs is of an "inverted U" shape (H3c). Adjective ratios of 9 % and adverb ratios of 5 % seem to be optimal. Since the regression analysis cannot provide answers to the question which adjectives and adverbs typically occur in helpful reviews, we supplemented a content analysis (Zhang and Tran's 2011). We used 10-fold cross validation to evaluate predictive accuracy and to obtain robust information scores. Reviews were divided randomly into 10 equally sized folds, of which we used 9 for training and 1 for evaluation. The training folds contained the 25 % most helpful and the 25 % most unhelpful reviews. Predictive accuracy – measured by precision (between 53.83 % and 73.82 %), recall (between 48.47 % and 77.87 %), and F-score (between 58.21 % and 71.64 %) – indicates good model fit for all categories (see Online-Appendix D).

Our analysis shows[6] that negative words (e.g., bad, disappointing) predominantly occur in unhelpful reviews. Positive words (e.g., fast, perfect) occur in both helpful and unhelpful reviews. Helpful reviews focus on product- and feature-related descriptions. Consumers of utilitarian goods in particular perceive information about delivery or customer service as very unhelpful (e.g., unfriendly, unanswered). Simple, easily comprehensible words (e.g., small, fast) seem to be most helpful, whereas complex or extremely positive words (e.g., ultra-light, exquisite) characterize unhelpful reviews.

To test H3d, we used the same regression equations as we did for analyzing

[5] We re-ran our regressions with other readability indices (Flesch-Kincaid Readability Ease, Flesch-Kincaid Grade Level, Gunning Fog Index, Automated Readability Index and Coleman-Liau Index) and found virtually no differences. We chose to present the model with SMOG because it produced the lowest values for the Akaike Information Criterion.

[6] The top 10 most (un)helpful adjectives and adverbs for each product category are listed in Online-Appendix D.

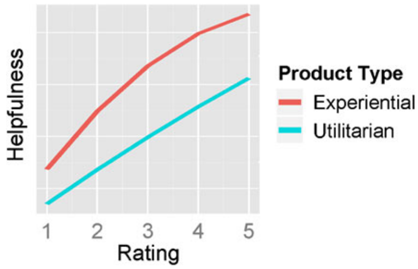

Fig. 3 Relationship between product rating and review helpfulness

H1a. The helpfulness of reviews for utilitarian products depends linearly on the products' rating. For experiential products' reviews, we found an "inverted U"-shaped relationship between rating and helpfulness. In the relevant interval between 1 and 5 stars, however, the relationship between rating and helpfulness for experiential products' reviews displays a monotonous increase (**Fig. 3**). H3d is rejected.

The relationship we identified between rating and helpfulness differs from findings by Ghose and Ipeirotis (2011), Mudambi and Schuff (2010), or Forman et al. (2008), but is also reported by Pan and Zhang (2011). All else being equal, reviews with positive product ratings are perceived as more helpful than reviews with negative product ratings.

5 Discussion

Structuring prior research based on Wang and Strong's (1996) multidimensional data quality framework has helped us develop one of the most comprehensive (Hao et al. 2009; Pan and Zhang 2011) helpfulness prediction models to date. We find that features from three dimensions – intrinsic, contextual and representational quality (Wang and Strong 1996) – determine review helpfulness.

Our data set contains all reviews from three utilitarian and three experiential product categories, making it, to the best of our knowledge, one of the most comprehensive sets to date.The results of our data analyses provide interesting implications for online retailers and researchers.

5.1 Implications for Practice

Online retailers can use our findings to (i) compute automatic helpfulness scores for reviews without votes and (ii) provide specific, individualized advice to reviewers. Helpful reviews possess six major characteristics.

1. Helpful reviews contain a moderate number of simple adjectives (at most 9 % of text) and adverbs (at most 5 %) (H3c) which focus on product features rather than service quality (Online-Appendix D).
2. Helpful reviews either discuss many product features or are the first to discuss a particular feature. In the latter case, they are perceived as helpful even if they do not discuss many features overall (H2b).
3. Dissenting opinions are perceived as particularly valuable. Reviewers need not fear they will be "punished" if their view does not conform to previously voiced opinions (H2a).
4. Reviews on experiential products are perceived as particularly helpful when reviewers describe their personal, subjective experiences rather than when they give an objective account of product features. The reverse is true for utilitarian goods (H1a).
5. Correct spelling improves review helpfulness; especially if a large number of reviews is available for (prospective) readers to choose from (H3b).
6. Reviews are not perceived as more helpful when reviewers reveal their true identity (H1c).

We suggest that online retailers implement our model in (i) an automatic review scoring system to help consumers find helpful reviews fast and (ii) a review writing tool which provides reviewers with instant personalized feedback during the writing process. Both measures will improve the average quality of the reviews on the retailing platform and make it easier for consumers to retrieve relevant information. This will improve the shopping experience and, in consequence, lead to increased customer satisfaction, retention times, and higher sales.

These advantages will only be realized fully, of course, if readers' acceptance of the automatic helpfulness scores and reviewers' acceptance of the writing guidelines is high. The degree of readers' acceptance of and trust in automatic helpfulness scores will probably depend greatly on how transparent the underlying mechanism is made. Future research on consumer trust in online reviews and online agents is required to shed light on this issue.

Reviewers' acceptance of the writing guideline will probably depend on whether they feel that it restricts their ability to express themselves (and on the degree to which their helpfulness scores actually improve). We suggest that the use of such a guideline be made entirely voluntary to avoid this pitfall. Developing a tool which implements our guidelines and balances the requirements of high review helpfulness and high individuality will be a very interesting challenge for future research.

5.2 Implications for Research

Our study provides two main contributions to research in eWOM which are of interest to both behavioral and design science-oriented BISE researchers. First, we adapted Wang and Strong's (1996) data quality framework to assess the effect of review quality factors on helpfulness. Having consolidated previous research within this framework, it is now easier to spot research gaps and explain contradictory findings. One possible reason for conflicting results by prior research is the fact that most studies examined variables from only one or two dimensions, although actually helpfulness appears to be determined by variables from three dimensions. A second potential explanation for these contradictions is the fact that many previous studies did not distinguish between product types, and a third explanation that previously used data sets may have been subject to a variety of biases.

Second, our empirical study shows that representational, contextual and intrinsic quality factors influence a review's perceived helpfulness significantly (see Sect. 4.2 for detailed findings). Reviews for utilitarian and experiential products display systematic differences: reviews for utilitarian products consist of more words, are harder to read, and are characterized by higher attribute entropy.

5.3 Limitations

In contrast to most prior research, we found that positive reviews exhibited better helpfulness scores than negative reviews. Unfortunately, our data does not contain observatory data on consumer perceptions and behavior, which might explain this phenomenon. We therefore encourage future research into the question of how consumers perceive positive and negative reviews and what motivates them to vote on a review.

Like most research on review helpfulness, ours rests on the assumption that helpfulness votes are unbiased and reflect perceptions of helpfulness truthfully. But not all consumers who read reviews also vote on them. We believe that a valuable contribution could be made by research into the question of whether systematic (and economically relevant) differences between true and revealed helpfulness exist.

Although our suppositions of how review writing guidelines and automatic scoring systems will affect customer satisfaction and purchasing behavior are backed by literature, they have never been tested empirically. We are currently implementing intelligent software which provides personalized feedback to reviewers and are planning to evaluate its effects on consumer attitude and behavior.

References

Benamara F, Cesarano C, Picariello A, Reforgiato D, Subrahmanian VS (2007) Sentiment analysis: adjectives and adverbs are better than adjectives alone. In: Proceedings of the 1st international AAAI conference on weblogs and social media, Boulder, CO

Berger J, Sorensen AT, Rasmussen SJ (2010) Positive effects of negative publicity: when negative reviews increase sales. Marketing Sci 29(5):815–827

Bone PF (1995) Word-of-mouth effects on short-term and long-term product judgments: interpersonal buyer behavior in marketing. J Bus Res 32(3):213–223

Burton J, Khammash M (2010) Why do people read reviews posted on consumer-opinion portals? J Mark Manag 26(3):230–255

Cacioppo JT, Petty RE (1984) The elaboration likelihood model of persuasion. Adv Consum Res 11:673–675

Chen CC, Tseng Y (2011) Quality evaluation of product reviews using an information quality framework. Decis Support Syst 50(4):755–768

Chen P, Dhanasobhon S, Smith MD (2008) All reviews are not created equal: the disaggregate impact of reviews and reviewers at Amazon.com. SSRN working paper. http://ssrn.com/abstract=918083. Accessed 2013-02-25

Connors L, Mudambi SM, Schuff D (2011) Is it the review or the reviewer? A multimethod approach to determine the antecedents of online review helpfulness. In: Proceedings of the 44th Hawaii international conference on systems science (HICSS), Hawaii, USA

Danescu-Niculescu-Mizil C, Kossinets G, Kleinberg JM, Lee L (2009) How opinions are received by online communities: a case study on Amazon.com helpfulness votes. In: Quemada J, León G, Maarek Y, Nejdl W (eds) Proceedings of the 18th international conference on world wide web. ACM, New York

Dellarocas C (2003) The digitization of word of mouth: promise and challenges of online feedback mechanisms. Manag Sci 49(10):1407–1424

Dellarocas C, Gao G, Narayan R, (2010) Are consumers more likely to contribute online reviews for hit or Niche products? J Man Inf Sys 27(2):127–157

Eagly AH (1974) Comprehensibility of persuasive arguments as a determinant of opinion change. J Pers Soc Psychol 29(6):758–773

Eagly AH, Chaiken S (1984) Cognitive theories of persuasion. In: Berkowitz L (ed) Advances in experimental social psychology. Academic Press, San Diego, pp 267–359

Feldman JM, Lynch JG (1988) Self-generated validity and other effects of measurement on belief, attitude, intention, and behavior. J Appl Psychol 73(3):421–435

Folkes VS (1988) Recent attribution research in consumer behavior: a review and new directions. J Consum Res 14(4):548–565

Ford GT, Smith DB, Swasy JL (1990) Consumer skepticism of advertising claims: testing hypotheses from economics of information. J Consum Res 16(4):433–441

Forman C, Ghose A, Wiesenfeld B (2008) Examining the relationship between reviews and sales: the role of reviewer identity disclosure in electronic markets. Inf Syst Res 19(3):291–313

Ghose A, Ipeirotis PG (2011) Estimating the helpfulness and economic impact of product reviews: mining text and reviewer characteristics. IEEE Trans Knowl Data Eng 23(10):1498–1512

Greene WH (2012) Econometric analysis, 7th edn. Prentice Hall, Upper Saddle River

Hao Y, Li Y, Zou P (2009) Why some online product reviews have no usefulness rating? In: Proceedings of the Pacific Asia conference on information systems (PACIS 2009), Hyderabad, India, Paper 100

Herr PM, Kardes FR, Kim J (1991) Effects of word-of-mouth and product-attribute information on persuasion: an accessibility-diagnosticity perspective. J Consum Res 17(4):454–462

Jain SP, Posavac SS (2001) Prepurchase attribute verifiability, source credibility, and persuasion. J Consum Psychol 11(3):169–180

Jin J, Liu Y (2010) How to interpret the helpfulness of online product reviews: bridging the needs between customers and designers. In: Cortizo JC, Carrero FM, Cantador I, Troyano JA, Rosso P (eds) Proceedings of the 2nd international workshop on search and mining user-generated contents. ACM, New York, pp 87–94

Kim S, Pantel P, Chklovski T, Pennacchiotti M (2006) Automatically assessing review helpfulness. In: Proceedings of the 2006 conference on empirical methods in natural language processing. Association for computational linguistics. Sydney, Australia, pp 423–430

Klare GR (2000) The measurement of readability: useful information for communicators. ACM J Comput Doc 24(3):11–25

Korfiatis N, Rodríguez D, Sicilia M (2008) The impact of readability on the usefulness of online product reviews: a case study on an online bookstore. In: Lytras MD, Carroll JM, Damiani E, Tennyson RD (eds) Emerging technologies and information systems for the knowledge society. Springer, Heidelberg, pp 423–432

Lahiri S, Mitra P, Lu X (2011) Informationality judgement at sentence level and experiments with formality score. In: Proceedings of the 12th international conference on

Abstract

Michael Scholz, Verena Dorner

The Recipe for the Perfect Review?

An Investigation into the Determinants of Review Helpfulness

Online product reviews, originally intended to reduce consumers' pre-purchase search and evaluation costs, have become so numerous that they are now themselves a source for information overload. To help consumers find high-quality reviews faster, review rankings based on consumers' evaluations of their helpfulness were introduced. But many reviews are never evaluated and never ranked. Moreover, current helpfulness-based systems provide little or no advice to reviewers on how to write more helpful reviews. Average review quality and consumer search costs could be much improved if these issues were solved. This requires identifying the determinants of review helpfulness, which we carry out based on an adaption of Wang and Strong's well-known data quality framework. Our empirical analysis shows that review helpfulness is influenced not only by single-review features but also by contextual factors expressing review value relative to all available reviews. Reviews for experiential goods differ systematically from reviews for utilitarian goods. Our findings, based on 27,104 reviews from Amazon.com across six product categories, form the basis for estimating preliminary helpfulness scores for unrated reviews and for developing interactive, personalized review writing support tools.

Keywords: Electronic commerce, Product reviews, Internet retailing, Electronic word-of-mouth

computational linguistics and intelligent text processing, Tokio

Li MX, Huang L, Tan CH, Wei KK (2011) Assessing the helpfulness of online product review: a progressive experimental approach. In: Seddon PB, Gregor S (eds) Proceedings of the Pacific Asia conference on information systems (PACIS 2011), Brisbane, Australia, Paper 111

Li X, Hitt LM (2008) Self-selection and information role of online product reviews. Inf Syst Res 19(4):456–474

Liu J, Cao Y, Lin C, Huang Y, Zhou M (2007) Low-quality product review detection in opinion summarization. In: Proceedings of the joint conference on empirical methods in natural language processing and computational natural language learning (EMN-CoNLL), pp 334–342

Liu Y, Huang X, An A, Yu X (2008) HelpMeter: a nonlinear model for predicting the helpfulness of online reviews. In: Proceedings of the 2008 IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology, Sydney, Australia, pp 793–796

McLaughlin HG (1969) SMOG grading – a new readability formula. J Read 12(8):639–646

Mudambi SM, Schuff D (2010) What makes a helpful online review? A study of customer reviews on Amazon.com. MIS Q 34(1):185–200

Nelson P (1970) Information and consumer behavior. J Polit Econ 78(20):311–329

Netzer O, Srinivasan V (2011) Adaptive self-explication of multiattribute preferences. J Mark Res 48(1):140–156

Otterbacher J (2008) Managing information in online product review communities: two approaches. In: Golden W, Acton T, Conboy K, van der Heijden H, Tuunainen VK (eds) Proceedings of the 16th European conference on information systems, pp 706–717

Pan Y, Zhang JQ (2011) Born unequal: a study of the helpfulness of user-generated product reviews. J Retail 87(4):598–612

Park DY, Lee J, Han I (2007) The effect of on-line consumer reviews on consumer purchasing intention: the moderating role of involvement. Int J Electron Commer 11(4):125–148

Schindler RM, Bickart B (2012) Perceived helpfulness of online consumer reviews: the role of message content and style. J Consum Behav 11:234–243

Schlosser AE (2011) Can including pros and cons increase the helpfulness and persuasiveness of online reviews? The interactive effects of ratings and arguments. J Consum Psychol 21(3):226–239

Scholz SW, Meissner M, Decker R (2010) Measuring consumer preferences for complex products: a compositional approach based on paired comparisons. J Mark Res 47(4):685–698

Schwenk CR (1986) Information, cognitive biases, and commitment to a course of action. Acad Manag Rev 11(2):298–310

Sen S, Lerman D (2007) Why are you telling me this? An examination into negative consumer reviews on the web. J Interact Mark 21(4):76–94

Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27(3):379–423

Toutanova K, Manning CD (2000) Enriching the knowledge sources used in a maximum entropy part-of-speech tagger. In: Proceedings of the joint SIGDAT conference on empirical methods in natural language processing and very large corpora, Hong Kong, China

Tversky A, Kahneman D (1974) Judgment under uncertainty: heuristics and biases. Science 185(4157):1124–1131

Wang B, Zhu W, Chen L (2011) Improving the Amazon review system by exploiting the credibility and time-decay of public reviews. Informatica 35(4):463–472

Wang RW, Strong DM (1996) Beyond accuracy: what data quality means to data consumers. J Manag Inf Syst 12(4):5–33

Weathers D, Sharma S, Wood SL (2007) Effects of online communication practices on consumer perceptions of performance uncertainty for search and experience goods. J Retail 83(4):393–401

Wright DB, London K (2009) Modern regression techniques using R. Sage, London

Wu P, van der Heijden H, Korfiatis N (2011) The influences of negativity and review quality on the helpfulness of online reviews. In: Galletta DF, Liang T (eds) Proceedings of the international conference on information systems, Shanghai, China

Xia L, Bechwati NN (2011) Word of mouse: the role of cognitive personalization in online consumer reviews. J Interactive Advertising 9(1):3–13

Xia R, Zong C, Li S (2011) Ensemble of feature sets and classification algorithms for sentiment classification. Inf Sci 181(6):1138–1152

Zhang JQ, Craciun G, Shin D (2010) When does electronic word-of-mouth matter? A study of consumer product reviews. J Bus Res 63(12):1336–1341

Zhang R, Tran T (2010) Helpful or unhelpful: a linear approach for ranking product reviews. Electron Commer Res 11(3):220–230

Zhang R, Tran T (2011) An information gain-based approach for recommending useful product reviews. Knowl Inf Syst 26(3):419–434

Zhang Z, Varadarajan B (2006) Utility scoring of product reviews. In: Yu PS, Tsotras V, Fox EA, Liu B (eds) Proceedings of the 15th ACM international conference on information and knowledge management, Arlington, VA. ACM, New York, p 51