Australasian Marketing Journal 21 (2013) 221–227

Contents lists available at ScienceDirect

Australasian Marketing Journal

journal homepage: www.elsevier.com/locate/amj

Some remarks on the internal consistency of online consumer reviews

Diana Schindler *, Reinhold Decker

Department of Business Administration and Economics, Bielefeld University, Postfach 10 01 31, D-33501 Bielefeld, Germany

ARTICLE INFO

Article history:
Available online 30 September 2013

Keywords:
Online consumer reviews
Internal consistency
Empirical analysis
eWOM

ABSTRACT

In recent years, online consumer reviews (OCRs) have developed into a popular topic of qualitative and quantitative marketing research. Depending on the platform they arise from, OCRs can be composed of an overall star rating, explicit pros and cons, free text comments and a recommendation indicator. This paper examines whether these components provide consistent information regarding the evaluated product. Insights about the internal consistency of this form of electronic word-of-mouth (eWOM) can be useful in consumer-oriented marketing strategies, complaint management as well as product development and improvement. By using a regression analysis framework the correlative relationships between star rating and recommendation are studied, as well as the effects that posted advantages and disadvantages, stated either explicitly or implicitly, have on these two types of product evaluations. The available results clearly support the view of OCRs as a consistent database for marketing research concerning eWOM. However, they also show that a combined consideration of all components can remarkably increase the information obtainable from this data source.

1. Introduction

Online opinion platforms, such as Amazon.com or Epinions.com are popular options for consumers to talk about individual experiences with products, to rate a particular product, and to recommend or not recommend it (Hennig-Thurau et al., 2004; Hennig-Thurau and Walsh, 2003). Research by both practitioners and academics characterises online posted opinions as a valuable means for the measurement of word-of-mouth (Bickart and Schindler, 2001; Dellarocas, 2003; Dellarocas et al., 2007; Chen et al., 2003; Chen and Xie, 2008) and an important source of information about aggregate product evaluation (Lee and Bradlow, 2011). From a marketing perspective, several advantages of online consumer reviews (OCRs) strongly motivate their use as an input for marketing decision-making, e.g. for product development and quality control (Dellarocas, 2003), automatic analysis of consumer preferences for product design and development purposes (Lee and Bradlow, 2011), or as an integral element of marketing communication (Chen and Xie, 2008). Furthermore, it is postulated that OCRs are a dynamic, daily-growing source of real-time information about consumers' needs and preferences. In contrast to traditional consumer surveys, consumers normally post their opinions without any external trigger or specification of topic. Consequently, OCRs can be assumed to be relatively authentic (Bickart and Schindler, 2001; Chen et al., 2003; Dellarocas, 2003; Dellarocas et al., 2007) and include notable aspects (Decker and Trusov, 2010; Ghose and Ipeirotis, 2007). Last but not least, they enable the identification of possibly unknown associations related to the products of interest. Positive associations that have been articulated by a considerable number of consumers can be used in two-sided advertising campaigns (Pechmann, 1992; Rucker et al., 2008), in which inevitable weaknesses or negative associations are qualified by references to regularly-discussed strengths and advantages. Insofar, the empirical findings of appropriate OCR analyses can also serve as an external reference point for one's own product communications.

1.1. Online consumer reviews for marketing research – related work

A closer look at current empirical OCR studies indicates that there are several concepts to analyse eWOM gathered from OCRs. The most frequently used components of OCRs are the overall star ratings and free texts (Tuma and Decker, 2013). With regard to the use of eWOM for marketing or business management, the former have particularly developed into a source of key information.

The paper by Chen et al. (2003) is one of the first to explicitly consider using OCRs for a holistic view of marketing strategies and to recommend OCR ratings as a valid source for eWOM analysis. Later, Dellarocas et al. (2007) follow this view and show the statistical significance of ratings for forecasting sales. With their study, they remove initial doubts (Duan et al., 2008) about the prediction quality of star ratings. However, there are also a remarkable number of studies that focus on the quantification or measurement of the impact of posted ratings on (future) product sales but also

* Corresponding author. Tel.: +49 5211064738; fax: +49 5211066456.
E-mail addresses: dschindler@uni-bielefeld.de (D. Schindler), rdecker@uni-bielefeld.de (R. Decker).

1441-3582/$ - see front matter © 2013 Australian and New Zealand Marketing Academy. Published by Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ausmj.2013.08.001

integrate other variables, such as, the volume of ratings, social interactions between reviewers, or the relationship between product and consumer characteristics (Chevalier and Mayzlin, 2006; Cui et al., 2012; Godes and Mayzlin, 2004; Moe and Trusov, 2011; Zhu and Zhang, 2010).

The studies, in particular, already imply that the sole use of rating information might result in limited insights. This finds further support in several studies: Hu et al. (2006), for example, state that the average star ratings do not represent aggregate preferences and, hence, the analysis of only ratings may result in wrong interpretations. Furthermore, Hu et al. (2009) empirically show that the distribution of star ratings for a specific product category often tends to be j-shaped. OCRs with low star rating values (1 or 2 stars) and with high rating values (5 stars or higher, depending on the scale used) predominate. From this finding, they conclude that aggregate star ratings – used as a sole database – cannot guarantee a consistent basis. Additionally, and compatible with the above-mentioned concerns, some other studies postulate that reviewers may be influenced by already-published OCRs and, hence, star ratings do not necessarily provide an objective quality criterion (Li and Hitt, 2008; Moe and Schweidel, 2012).

Apart from the growing number of star-rating-based studies, several researchers deal with the analysis of free text comments and the inherent sentiment values. Hu and Liu (2004), for example, illustrated several techniques to automatically mine features that have been commented on by reviewers (see also Liu et al., 2005). In both studies, the main purpose was to provide user decision support systems, which assist consumers in collecting information about products that they want to buy. Apart from this, various approaches have been suggested in combination with product development. For example, Gamon et al. (2005) and Kim et al. (2009) discuss techniques to support manufacturers by automatically evaluating features that should be improved. With an explicit focus on OCRs as an input or starting point of standard marketing research, Lee and Bradlow (2011) recently used an automatic evaluation technique to generate attributes from pros and cons that can be used in conjoint analysis. In a similar context, Decker and Trusov (2010) demonstrate ways in which OCRs can be used in a preference measurement setting and how the resulting market knowledge can be used for data-driven product development or improvement processes.

1.2. Implications for research

According to the above literature analysis, OCRs constitute a database of increasing importance to explore consumer feedback and to observe eWOM activities. However, although (overall) star ratings, as a main component of OCRs, have developed into a standard input for marketing related eWOM analysis, several researchers have doubts about their general reliability. For example, from a consumer's perspective, the credibility of an OCR can be assumed to be higher when internal consistency is evident. Tsang and Prendergast (2009) even show that inconsistent text-rating reviews (e.g. negative text with a positive rating) may lead to a decline in trust. Accordingly, with respect to trustworthiness and reliability, analyses of OCRs have to take into account internal consistency. Therefore, we assume all components of an OCR to be interdependent and contribute to the overall picture.

As already indicated above, the most typical component of an OCR is the numerical overall star rating (referred to as the "star rating" in the following). Furthermore, there may exist explicitly stated advantages (referred to as "pros") and disadvantages (referred to as "cons") in the "pros & cons panel", information about the recommendation or non-recommendation of the product, and a free text comment of varying length (referred to as "free text"). Last but not least, OCRs are often completed by the selection of predefined product attributes that are rated by the consumers. The resulting attribute-wise scores are not focused on in this study since the meaningfulness of corresponding analyses strongly depends on the particular set of attributes that are considered. An example of a corresponding OCR is illustrated in Fig. 1.

The above discussion implies that the empirical investigation of the internal consistency of OCRs is a basic challenge and a prerequisite for a sustainable establishment of this comparatively new field of marketing research in practice. Therefore, the current paper examines the question of whether or not plausible interrelationships can be detected between different components of an OCR and if these components constitute an internally consistent database for further comprehensive analyses. Insofar, the following is also a discussion of data quality in the OCR context. The relationships of interests are illustrated by Fig. 2.

The advantages (the "pros") and disadvantages (the "cons") of a product can either be articulated in the pros & cons panel of the OCR (see Fig. 1) or they can be expressed in the free text. The sentiment values (positive/negative) of the attributes occurring in the pros & cons panel are explicitly given by their categorisation. We therefore refer to this data as the *explicit* pros and cons of a product. In contrast to this, the sentiment values of the product attributes that occur in the free text have to be determined using appropriate opinion mining methods. The attributes that, positively or negatively, characterise the product are referred to as *implicit* pros and cons. In Fig. 2, both components are represented by the label "advantages and disadvantages." The interesting relationships between the individual OCR components and the respective directions are illustrated by means of arrows. We will refer to these relationships in Section 3 when specifying corresponding research hypotheses.

Subsequently, we first give a brief description of the necessary data pre-processing steps. Then, the research hypotheses and the methodological framework to verify them are presented. The main part of the paper will focus on empirical results on internal consistency, assessed on the basis of the significance and plausibility of the parameters of problem-oriented regressions. Finally, we discuss the available results, outline limitations, and point to possible directions of future research.

2. Data pre-processing and the resulting data structure

The database used in this study comprises ca. 17 k OCRs about digital cameras. The pre-processing of the explicit pros and cons is based on a procedure similar to that described in Decker and Trusov (2010). Accordingly, in the first step, the comments in the pros & cons panel of the OCRs are divided into individual words and phrases. Those that neither represent a directly-mentioned attribute (e.g. "reliability") nor an indirectly mentioned attribute (e.g. "expensive", corresponding to "price") are eliminated. Then, redundant terms are aggregated and synonyms are merged. To further improve the data quality the list of attributes is matched with product descriptions of leading manufactures. Finally, the data is dichotomised: We therefore define a binary variable a_{ij}^{proEX} which represents the occurrence (coded as 1) of attribute j ($j = 1, \ldots, J$) in the *pros*-field of the pros & cons panel of the OCR i (referred to as an *explicit pro*). Analogously, a_{ij}^{conEX} denotes the occurrence (coded as 1) of attribute a in the *cons*-field of the panel (referred to as an *explicit con*). The non-occurrence is coded as 0.

In order to identify the attributes that are discussed by the reviewers in the free texts, we made use of a lexicon-based approach (Ding and Liu, 2008). The free texts are matched with the list of attributes available from the procedure described in the previous paragraph. Then, the sentiment values of the attributes are determined. This is done by (1) measuring the within-sentence