EXPLORE AND CLASSIFY USER GENERATED PRODUCT REVIEWS FROM THE INTERNET: A CASE STUDY

Jasper Pieterse, Ibai Perez Artieda, Aylin Koca and Lu Yuan
Department of Industrial Design
Eindhoven University of Technology, The Netherlanda
t.j.pieterse@student.tue.nl; i.perez.artieda@student.tue.nl; a.koca@tue.nl; y.lu@tue.nl

ABSTRACT

The overall aim of this paper is to demonstrate how to obtain reliable information from the user generated product reviews on the Internet. Most of these reviews only contain information about products' soft failures, which result in the rejection of innovative consumer electronics products that are not due to broken hardware or software, but due to "broken expectations" of users. However, the reliability of this essentially subjective information is difficult to evaluate. In order to evaluate the reliability of the user generated product reviews, we explore a method to first classify them using the Disconfirmed Expectations Ontology (DEO) that is developed to analyze soft reliability issues, and then further to explore their information content by means of user tests. We apply this method in a case study using a consumer electronic product from a multi-national company.

INTRODUCTION

Products in the consumer electronics market are becoming more innovative and complex. This has led to a shift in the trends for new product development in the past decades (Ouden, 2006). It has become harder for product developers to maintain product quality and reliability. As developers aim to maintain a high level of quality, from a consumer perspective, issues other than the classical technical product quality are playing an important role; namely the expectations that consumers have on the product and their personal experiences with these products. These expectations and experiences are predominantly driven by how the product functions and features are used and experienced in the daily lives of its users; i.e., the so-called soft reliability of the product. Especially consumer complaints are currently showing a rise regarding soft reliability issues, which result in the rejection of innovative consumer electronics products, not due to broken hardware or software, but due to 'broken expectations' of users. Misalignments between product capabilities and consumers expectations or experiences affect the overall success of a product in the marketplace. In recent years, it has been observed that misalignments increasingly lead to rejection of consumer electronic products that are fully functional according to their technical specifications (Ouden, Lu ,Sonnemans, Brombacher, 2006). On the other hand, the quality monitoring systems of manufacturing companies are not tuned to capture these soft reliability issues, despite their various data resources available for obtaining rich qualitative user feedback from the field. Addressing these issues is crucial to better align real contextual needs and preferences of users with product capabilities. In this paper, we focus on the Internet as a source to obtain rich feedback from actual users as it proves to be a naturally emerging source of user feedback data, where users exchange rich qualitative accounts of products -both positive and negative. As product designers and developers are starting to recognize the wealth of

information available through this medium, it is still underutilized (Koca, Brombacher, 2008). It could give product developers an opportunity to analyze and use this information to better understand their products' strengths and weaknesses. However, this essentially subjective information, as well as its reliability, is difficult to evaluate. Therefore, we explore a method to first classify the user generated product reviews using (DEO) (Figure 1), and then further explore the most prominent product feedback (positive and negative) by means of user tests.

DISCONFIRMED EXPECTATIONS ONTOLOGY

DEO is a user-centered failure classification model that is being developed at Eindhoven University of Technology (Koca, Lu, Brombacher, Hartmann, 2006) (Koca, Schouwenaar, Brombacher, 2007) (Koca, Schouwenaar, Brombacher, 2007). It is being developed as a generic and extendible tool to analyze different sources of user feedback data. DEO is used to capture negative and positive soft reliability issues from user feedback. Each soft reliability issue can be divided into one of seven smaller groups of problems (negative) and marvels (positive), resulting in fourteen different types of disconfirmed expectations that are defined as shown in figure 1. The resulting data can be prioritized based on occurrence frequency and communicated to relevant stakeholders within the product development (Koca, Karapanos, Brombacher, 2009)] The first four categories indicates where in the "use-phase" the users may have problems or marvels, they are time dependent, while the other three have no relation to any consumer use phase. Each individual positive category also corresponds to a relative negative category, for example a "Wow factor" in the product success could be interpreted as the opposite of the missing feature category in the product concept soft failures.

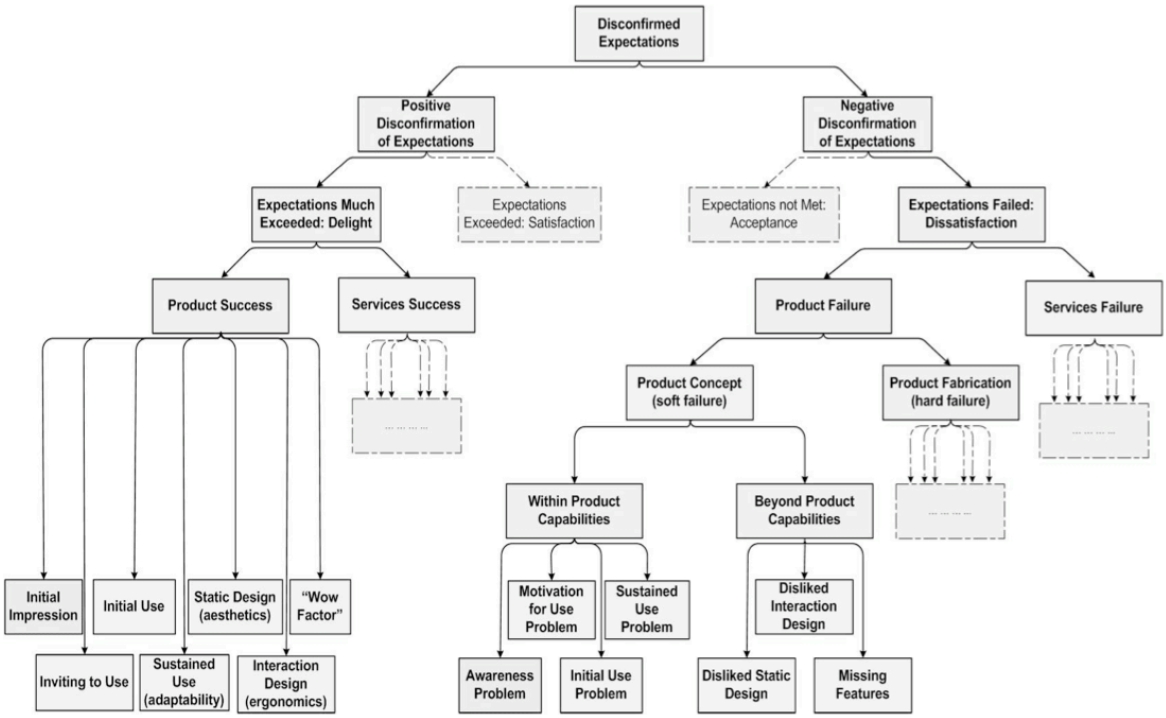

Figure 1; Disconfirmed Expectations Ontology (DEO) is for capturing negative and positive soft reliability issues from detailed user feedback data.

METHOD

We propose the following method to analyze user feedback from the Internet to obtain more valuable information about product soft reliability issues. The first step is to collect user generated product review from the internet. The data of these product reviews can be freely compiled by the user, they publish their own experiences without having to conform to any codification or classification scheme (Koca, Brombacher, 2008) resulting in mostly unstructured and difficult to (automatically) process data. Therefore each review is individually analyzed and statements regarding the product are extracted. The statements containing soft reliability issues are then classified using DEO. These DEO classifications give an indication of which product features contribute to the success of the product and whether the experienced problems are within or beyond the product's current capabilities. After the DEO classifications the individual statements are then coded by content, statements with a similar positive or negative experience are giving the same code. This creates groups of statements and allows for a ranking based on the frequency of occurrence. In order to further explore the reliability of the collected grouped statements, a qualitative user test is performed. User groups, fitting the product's target group, are given the product for one week to experience its features. Afterwards they are asked to what extent they experienced the statements collected earlier on from the Internet (i.e., regarding strengths and weaknesses) and also to rate their relative importance. The users are asked to rate the statements from 1 to 5 indicating if they have experienced the statement, after that they are asked to rate the statements importance from 0 to 5.

Going further into the actual statements, it is possible to see to what degree they are experienced and how important they are according to the consumer. This outcome can be used by product development to better understand how their product is being perceived and what features need to be kept, could be improved, or should considered to be added to the product.

Case study

The proposed methodology has been applied in the analysis of a consumer electronic product from a multi-national company. This product is a new generation of an existing product. It is a product for everyday use. The target group consists of consumers who own the previous product generation. Around 300 user generated reviews of this product have been analyzed from the following sited; Amazon.de, ciao.fr, bestlist.nl, kieskeurig.nl and wehkamp.nl resulting in close to 700 product statements. Figure 2 and 3 show the overall classifications and their frequency of occurrence, indicating which product feature contributes to the success of the product and if the experienced problems are within or beyond the product's current capabilities

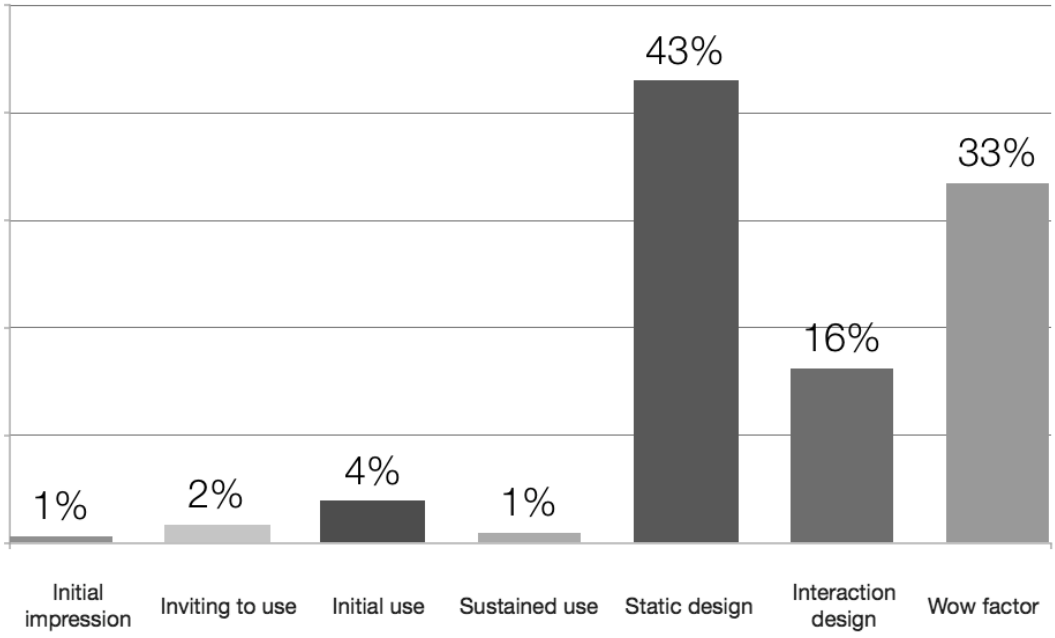

Figure 2: DEO classification product success

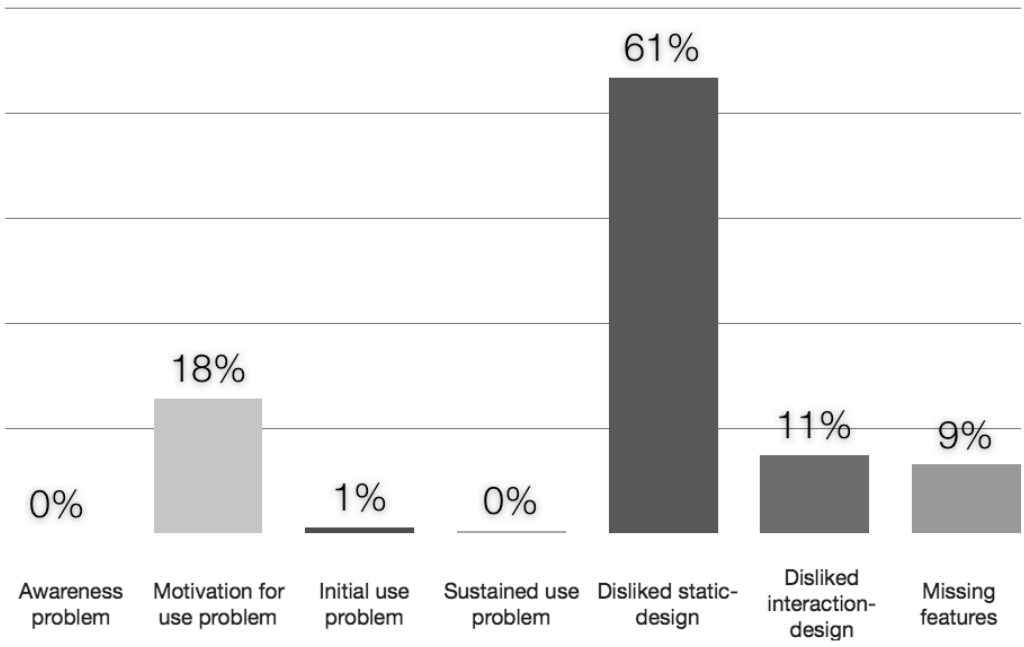

Figure 3: DEO classification soft failure

As apparent from the figures most of the categories, in both the positive and negative disconfirmations are not time related and thus giving no indication of use-phase. This is due to the fact that the information from the individual internet user generated product reviews rarely has a use-phase indication which is needed to classify them into time related categories. This means that the statements were classified into the latter categories only giving an indication where the statement occurs (looking at the product) but not when with regards to the use phase. Therefore no conclusion can be drawn about the time-effects in the product adoption during prolonged use. Another reason might be that the product capabilities were recognized by the users resulting in low frequency of occurrence of soft failures in the

of early use phases. From the positive disconfirmations (figure 2) it becomes apparent that the design of this product is valued by the users. They appreciate its features and how they are implemented (43% Static design). The users are also pleasantly surprised by the features of the product they did not envision (33% wow factor) and content with interactivity of its features (16% interaction design). Regarding the negative disconfirmations (figure 3), by far most users dislike the way the product features are designed (61% Disliked static-design) or they miss them completely (9% Missing features). There also seems to be a low motivation for using functions that the users are aware of but don't want to use (18% Motivation for use). Looking at the total results, the design of the product is the most conflicting criteria, 43% positive to 61% negative. In addition there is appears to be a large amount of features that are strongly pleasing which the user would never have envisaged (33% wow factor).
Only statements which occurred more than five times were grouped based on their frequency of occurrence and further evaluated in the user test. The objective of the user test was to gain insight about the relation between the statements that the user experienced in the user test and the importance that they considered this experience. A qualitative user test is performed with nine users who owned the previous generation of the product. They were given the product for one week due to time constraint to experience its features. Afterwards they were given a questionnaire in which are asked to rate from 1 to 5 according to the extent they agreed with the statements, after that they were asked to rate the statements importance from 0 to 5.

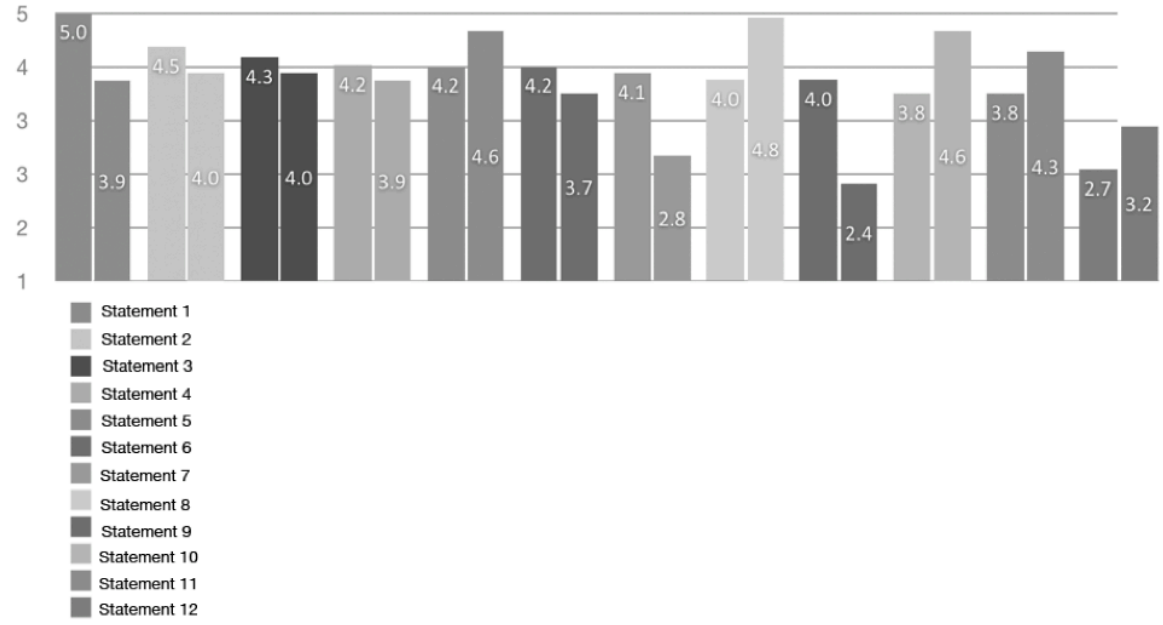

Figure 4: Average user importance (left bar of each pair) and experienced (right bar of each pair) of the positive statements

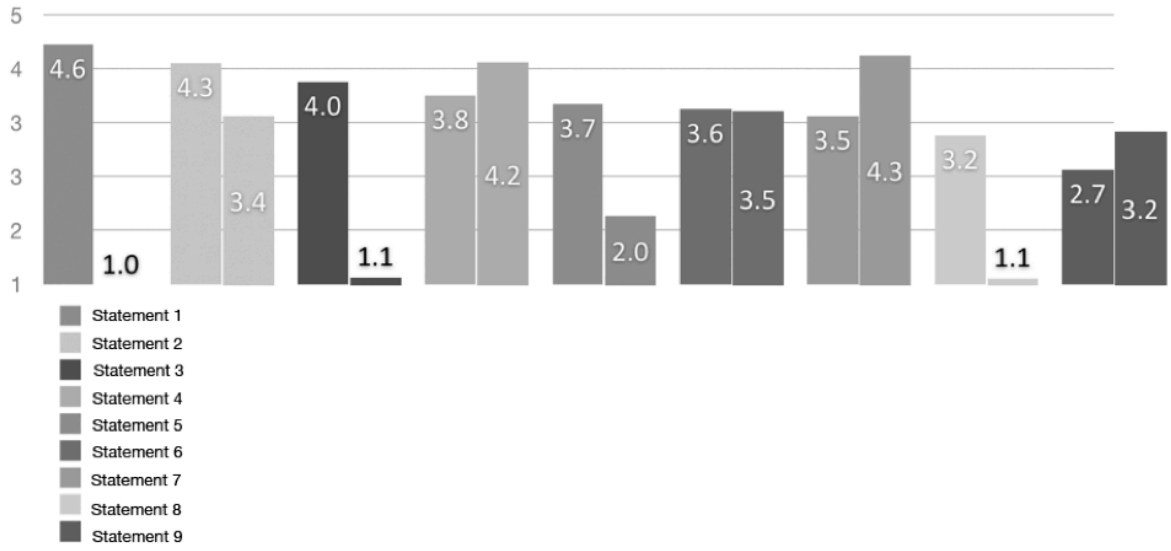

Figure 5 Average user importance (left bar of each pair) and experienced (right bar of each pair) of the negative statements

Among the positive statements, the experience and importance of two statements were contradictory. They were related to design and ease of cleaning. These two statements were not agreed, but they were considered as important. Based on this information, one can conclude that there could be the case of missing features or unrealized product promises. Necessary product improvement actions can be taken to solve this situation. The rest of the statements were both agreed and of high importance, indicating that the subjects of the statements were both important for the product and positively implemented.

Most of the negative statements presented to the users were considered as important even though they were not experienced by the users. This could be due to the following two reasons: the test period was too short to experience these statements, the test participates did not represent the users from the internet or the product performed really well. More research is needed here to be able to explain this situation.

The results were discussed with the company. They were very surprised by the obtained results as 75% of the results matched with the analysis of the usually used product performance data sources such as development test, help-desk, service center and etc. They would like to take further actions to explore the use of internet data in product development.

Conclusion and discussion

We proposed a method to evaluate the reliability of internet data by combining the use of DEO method and user test. This method was first evaluated with a consumer electronic product. The test results and the reaction from the company suggest that the potential of internet data in support product development is evident.

Yet, we need to mention a number of factors that may contribute to the limited results we have obtained. The online user generated product reviews used in the case study were collected from sites from three different countries. Using different countries might give a distorted view as cultural differences and product launch time are not taken into account. Looking back at our proposed method, we see that by applying DEO on the Internet data, almost no use-phase related data could be extracted. This may due to 1) relative low complexity of the product or 2) it is simply not part of the user generated product review. In

the first case it is positive for the product, in the latter case a part of the power of the DEO classification about the use-phase is missed that might provide useful information. This should then be addressed when developing information collection sites as it could increase the usefulness of our method. Furthermore, each of the statements has been analyzed and then evaluated by us, which may have introduced bias. Even though raters' bias is inherent to post-processing unstructured product review data on the Internet, a more structured approach, e.g., by embedding DEO-classes to be used by users as they submit their reviews on the Internet, can eliminate post-processing by dedicated raters and increase the reliability of the analysis. Finally the amount of users and the duration of the test were limited. In further exploration the duration and amount of users of the test should be increased to get more reliable data. The week the users now had is barely sufficient to experience all the product features, if a more complex product is used the users test should be prolonged

Managerial Significance

From a managerial perspective, this research can be used to develop an Internet-based product review system to collect user-generated product usage feedback, which automates the information extraction (e.g., via DEO), which can then be used as basis for a more detailed user testing. Utilizing the rich information continuously coming from the market in this way, product developers and managers can quickly gain an advantageous position against competition, by closely monitoring how their product is being perceived, and by adapting their product to better match with prominent user likes and dislikes. Furthermore, for user-generated product review sites, like Amazon, a new business opportunity can be envisioned through standardizing the information collection to harvest and sell it to product developers as useful data.

Koca, A., Brombacher, A.C. (2008a) Extracting "Broken Expectations" from Call Center Records: Why and How. In Proceedings of CHI 2008 Extended Abstracts on Human Factors in Computing Systems, Florence, Italy, 2008, pp. 2985-2990, ACM.

Koca, A.; Brombacher, A.C. (2008b) User-Centered Analysis of Feedback Operations for Quality Improvement in New Product Development. In Proceedings of the 15th International Annual EurOMA Conference, Groningen, The Netherlands, EIASM.

Koca, A.; Karapanos, E.; Brombacher, A.C. (2009b) "Broken Expectations" from a Global Business Perspective. In Proceedings of CHI 2009 Extended Abstracts on Human Factors in Computing Systems, Boston, USA, 2009, pp. 4267-4272, ACM.

Koca, A., Lu, Y., Brombacher, A.C. and Hartmann, J.H. (2006) Towards Establishing Foundations for New Classes of Reliably Problems Concerning Strongly Innovative Products. *Proc. ICMIT 2006. IEEE, 581-584.*

Koca, A., Schouwenaar, A.J.M. and Brombacher, A.C. (2007) Analysis of User-|Centered Failure Mechanisms in New Product DEvelopment for Quality Improvement. *Proc. EurOMA 2007*

Koca, A., Schouwenaar, A.J.M. and Brombacher, A.C (2007). Field-Feedback in Innovative Product Development: A Comparison of Two Industrial Approaches. *Proc IPDM 2007, 657-668*

Ouden E.d. (2006) Development of a design analysis model for consumer complaints : revealing a new class of quality failures, Eindhoven: Technische Universiteit Eindhoven

Ouden, E. d., Lu, Y., Sonnemans, P.J.M., and Brombacher, A.C. (2006) Quality and Reliability Problems from a Consumer's Perspective: An Increasing Problem Overlooked by Businesses? Quality and Reliability Engineering International 22, 7, 821-838.