Title	Harnessing the Experience Web to Support User-Generated Product Reviews
Authors(s)	Dong, Ruihai, Schaal, Markus, O'Mahony, Michael P., McCarthy, Kevin, Smyth, Barry
Publication date	2012-09-03
Publication information	Dong, Ruihai, Markus Schaal, Michael P. O’Mahony, Kevin McCarthy, and Barry Smyth. “Harnessing the Experience Web to Support User-Generated Product Reviews.” Springer, September 3, 2012. https://doi.org/10.1007/978-3-642-32986-9_7.
Conference details	Proceedings of the 20th International Conference, ICCBR 2012, Lyon, France, September 3-6, 2012.
Publisher	Springer
Item record/more information	http://hdl.handle.net/10197/4343
Publisher's statement	The final publication is available at www.springerlink.com
Publisher's version (DOI)	10.1007/978-3-642-32986-9_7

Harnessing the Experience Web to Support User-Generated Product Reviews

Ruihai Dong, Markus Schaal, Michael P. O'Mahony, Kevin McCarthy, and Barry Smyth

CLARITY: Centre for Sensor Web Technologies
School of Computer Science and Informatics
University College Dublin, Ireland
firstname.lastname@ucd.ie
http://www.clarity-centre.org

Abstract. Today, online reviews for products and services have become an important class of user-generated content and they play a valuable role for countless online businesses by helping to convert casual browsers into informed and satisfied buyers. In many respects, the content of user reviews is every bit as important as the catalog content that describes a given product or service. As users gravitate towards sites that offer insightful and objective reviews, the ability to source helpful reviews from a community of users is increasingly important. In this work we describe the Reviewer's Assistant, a case-based reasoning inspired recommender system designed to help people to write more helpful reviews on sites such as Amazon and TripAdvisor. In particular, we describe two approaches to helping users during the review writing process and evaluate each as part of a blind live-user study. Our results point to high levels of user satisfaction and improved review quality compared to a control-set of Amazon reviews.

1 Introduction

Among the many varieties of user-generated content on the social web, user-generated product reviews have come to play a vital role in a wide range of product/service oriented settings. Today, leading online product and service providers, from Amazon to TripAdvisor and iTunes to Etzy, prioritise user-generated reviews alongside catalog content in a variety of ways. For example, Amazon prioritises such reviews by placing the average review score directly beneath a product title. Moreover, it provides interested users with one-click access to review content, which includes recommendations of the most helpful, favourable and critical reviews. The rationale for such access is that review information of this kind has been shown to play a valuable role when it comes to helping users to make purchasing decisions, increasing the conversion rate of casual browsers into informed, satisfied buyers. For example, Hu et al. [10] describe the results of one study on the value of online reviews, concluding that consumers do understand the value difference between favourable and unfavourable opinions and respond accordingly. Furthermore, when consumers read online reviews, they pay attention not only to review scores, but to other contextual information such as a reviewer's reputation and reviewer exposure. The market responds more favourably to reviews written by reviewers with better reputation and higher exposure. In related work, Zhu et al. [16] examine the influence of online reviews on video game sales, indicating that reviews are more influential for less popular games.

While many sites prioritise the promotion of user-generated review content, they do little to support users when it comes to generating reviews, beyond the provision of a simple text-input review-form. At best, this provides friction when it comes to attracting new reviews from first-time reviewers, but further it may negatively impact the quality of the reviews that are provided. For this reason researchers have begun to consider ways in which users can be better supported during the review writing process. For example, the seminal work of Bridge et al. [3] describes a system called GhostWriter, which is designed with this task in mind. Briefly, GhostWriter uses case-based reasoning techniques to harness a collection of past review experiences, which are then used as a source of suggestions for reviewers as they write. Essentially, GhostWriter suggests fragments of similar reviews as hints to the reviewer.

In this paper, we build on the GhostWriter idea and extend it in a number of important ways. Firstly, we use association rule mining techniques to extract correlated product features from raw review experiences. These features correspond to recurring review fragments across a collection of reviews; for example, we might notice that digital camera reviews which refer to *picture quality* and *color saturation* also refer to *white balance*; see also [8]. Secondly, we consider a new topic-oriented approach that allows us to map raw review fragments to more structured product topics, and so avoid recommendation redundancy and improve the breadth of coverage of suggestions. Thirdly, we describe how these ideas have been implemented in the *Reviewer's Assistant* (RA) in the form of a browser plugin, which facilitates direct integration with sites like Amazon, TripAdvisor, etc. Finally, this paper also includes a comprehensive live-user trial of RA, in which we evaluate the perceptions of end-users and the quality of the reviews they produce; ultimately demonstrating how use of the RA can lead to higher quality reviews than those currently found on Amazon.com, at least in terms of the evaluation domain of digital cameras.

2 Related Work

Recent research indicates that online product reviews have a significant influence on the purchasing behaviour of users; see [10, 16]. For example, the effect of consumer reviews on book sales on Amazon.com and Barnesandnoble.com has been investigated by Chevalier & Mayzlin [5], concluding that the relative sales of books on a site correlates closely with positive review sentiment; although interestingly, there was insufficient evidence to conclude that retailers themselves benefit from making product reviews available to consumers. Similarly, Dhar & Chang [7] found a correlation between the volume of blog posts about a music album and future sales. Likewise, it has been shown that the early volume of online movie reviews can be used as a proxy for early sales [6].

Given the importance of product reviews it is not surprising that retailers and researchers have started to study different ways to help interested users find high quality reviews for products they are considering. As the volume of user-generated review content grows, it will become increasingly important to rank reviews based on some measure of relevance and/or helpfulness. For example, the work of [11–14] each consider different factors such as reviewer reputation, product genre familiarity, and review recency to automatically rank reviews based on their predicted helpfulness. Indeed, as the importance of online reviews has increased so has the temptation for interested parties to manipulate reviews to generate a bias for or against particular products. This in turn has motivated researchers to consider ways in which suspicious reviews can be identified and eliminated; see for example the work of Wu et al. [15].

In this work, we approach the problem of review quality from a different but complementary perspective: rather than attempt to rank reviews by helpfulness, or eliminate biased reviews, we focus on the process of writing a review in the first place. Our aim is to not only produce a better quality of user-generated reviews, but also to increase the number and diversity of reviews by attracting first-time reviewers who might be initially daunted at the prospect of writing a product review. Our work is inspired by the GhostWriter system, first introduced by Bridge et al. [4], as an approach to supporting users to create online adverts for personal goods and items they wish to dispose of, by using conversational CBR techniques [2] to reuse fragments of adverts of previously posted items. Later Healy and Bridge [9] adapted this approach to support users during the generation of product reviews, again using conversational CBR techniques, but this time recommending review fragments from past reviews that are similar to the user's current, incomplete review. Currently GhostWriter 2.0 [9] extracts noun phrases from these past product reviews (cases), and suggests these phrases directly to the reviewer/user.

Our Reviewer's Assistant is a close relative of the GhostWriter systems. It too harnesses past review experiences to proactively suggest review topics as the user writes their review. In this paper, we will describe how the RA combines association rule mining and topic extraction techniques with conversational CBR to generate these recommendations. In addition, the RA has been designed to fully integrate with existing online services such as Amazon and TripAdvisor, allowing users to use existing review tools while benefitting from suggestions made by the system.

3 The Reviewer's Assistant

The Reviewer's Assistant has been implemented as a browser plugin so that it can seamlessly integrate with pre-existing services like Amazon and TripAdvisor at the user interface level, providing support to users as they write their reviews, but without the need for backend integration with these underlying services. In this section we will describe the basic system architecture and key technical features of the RA system and provide an example of the RA in operation.

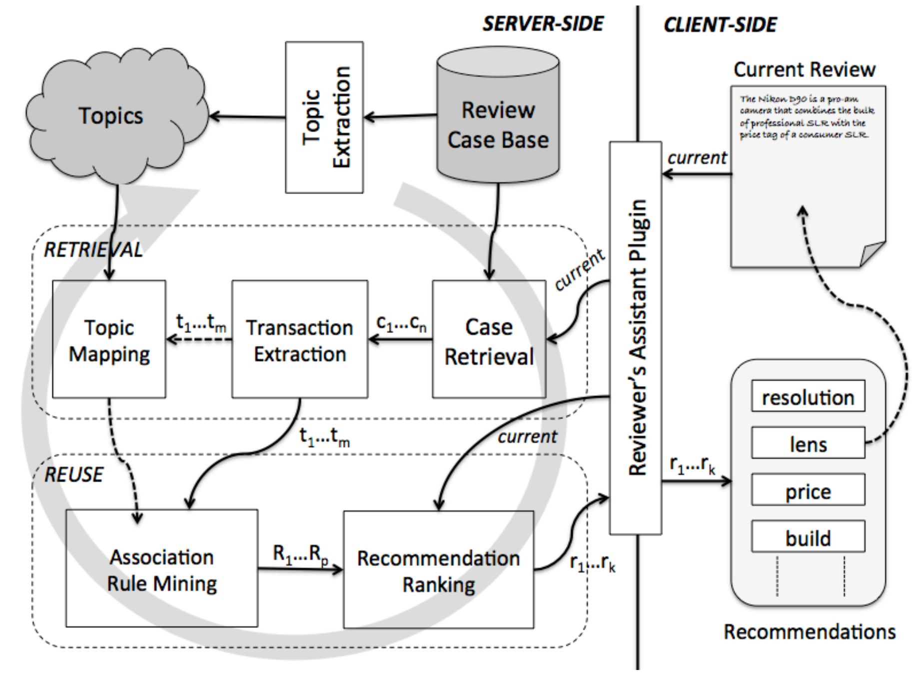


Fig. 1: The Reviewer's Assistance System Architecture.

Figure 1 provides a high-level overview of the Reviewer's Assistant system. The basic recommendation cycle is triggered on an ongoing basis as the user writes their review. During each recommendation cycle the user's current review text is used to identify a set of similar review cases. These cases are mined to extract a set of term-based association rules. These rules are then used to map the current review text to a ranked list of concrete topic recommendations for the user.

3.1 Case Discovery and Topic Extraction

Before describing the technical details of the basic RA recommendation cycle we will first look at the structure and source of case and domain knowledge.

The Review Case-Base. The RA is designed to operate over specific product domains and maintains a separate review case-base for each. A domain is assumed to be made up of a collection of products that share similar features, which ultimately will act as possible review targets for the reviewer. Each case corresponds to a previous review and includes the product id, the review text, and any meta information available, such as the overall review score or helpfulness. These reviews are automatically extracted from the underlying service by using any available API to extract relevant product and review data; typically the RA will extract only high-quality reviews (based on any helpfulness/quality meta data that is available).

Topic Extraction. In this paper we consider a version of the RA that maps review terms to product topics in order to improve recommendation quality. For the purpose of this feasibility study we adopt a very simple approach to topic modeling based around a hand-coded set of topics for the target domain, with each topic associated with a synonym set. As such the topic extraction component of Figure 1 is not currently automated but rather left as a matter for future work – assuming the value of our topic-based variation can be demonstrated in live usage.

3.2 Retrieving Similar Reviews

As the user writes their review, the review text is periodically (typically on the completion of a new sentence) used as query *current* against the relevant domain case-base to retrieve a set of similar reviews, from which term-based transactions are extracted as the basis for association rule mining.

Case Retrieval. In the current implementation we rely on a simple term-based Jaccard similarity metric to retrieve a set of the n review cases that are most similar to *current*. At present, this retrieval process is further restricted to only consider review cases that match the target review product id. For example, if the user is reviewing the latest "Nikon D90" camera, then only past reviews of this product will be considered for retrieval. Obviously this condition could be relaxed to facilitate the consideration of similar products but this is left as a matter for future work.

Extracting Transactions. At this point each of the n retrieved reviews is converted into a set of sentence-level transactions and a review-level transaction. This is a straightforward process that starts by identifying the nouns in a review text and then converts each sentence or review into a set of these nouns based on their order of appearance; see Figure 2. In the next subsection, we will describe how these transaction-based representations can be used by association rule mining techniques as a basis for recommendation.

Topic Mapping. In one version of the RA, *non-topic*, recommendation proceeds based on the mining of these noun-based transactional representations. However, in this paper we also consider a topic-based approach (we refer to this simply as *topic*) which first maps the raw transaction terms onto the topics

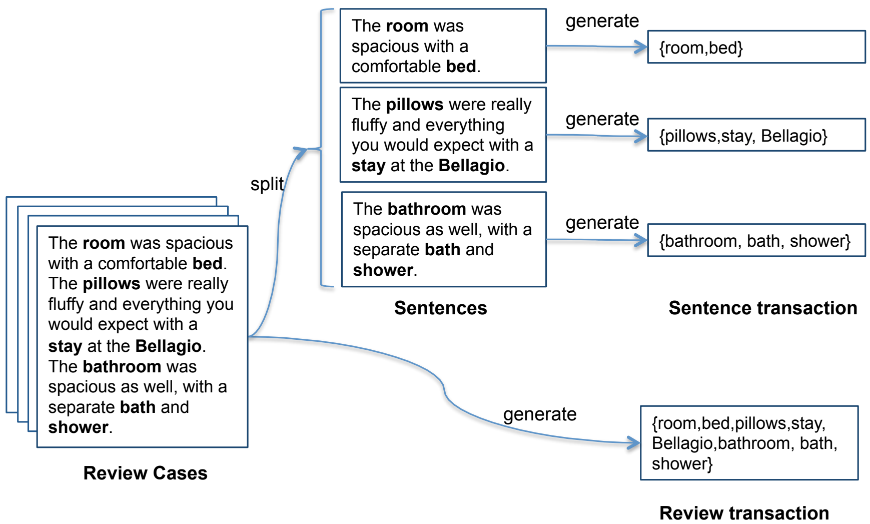


Fig. 2: Preparing Transactions for Association Rule Mining (ARM).

extracted from a given review collection as mentioned above. The purpose of this approach is that it affords a level of abstraction (topics vs. nouns) that has the potential to provide a more intuitive set of recommendations based on more meaningful product topics, rather than on the looser vocabulary of user generated reviews. As part of our evaluation in Section 4 we will consider whether this topic variation in fact translates into any meaningful evaluation benefit.

3.3 Generating Recommendations

The RA generates a set of ranked recommendations by using association rule mining techniques to discover patterns of nouns/topics that recur frequently across many reviews. In the following section we will describe how these rules can then be applied to the current review text in order to produce a ranked set of noun/topic recommendations to be returned to the user via the RA plugin.

Rule Mining. At this point we have a set of transactions (whether *non-topic* or *topic*), which reflect frequent collections of nouns/topics that occur at the sentence-level or review-level. For example, in the digital camera domain we might have transactions such as $\{$*image, lens, resolution*$\}$ and $\{$*size, price*$\}$ extracted at the sentence-level to indicate that review sentences discussed camera resolution, lens type and image quality or camera size and price. We can apply association rule mining [1] to identify frequently occurring transactions and to generate a set of association rules of the form $\{image, lens\} \rightarrow \{resolution\}$. Following the standard algorithm for association rule mining, we first filter-out rules that fall below a minimum *support* level; that is we keep subsets of transactions that have a pre-defined frequency of occurrence as candidates for rules and their antecedents, so-called itemsets.

Ranking Recommendations. The resulting rules are ranked in descending order of their confidence, which is basically an estimate of the probability of finding the topic/noun that forms the rule consequent given the occurrence of the antecedent. To generate a set of ranked recommendations we apply each of the extracted rules, in order of confidence, to the current review text. If the current review text triggers a rule of the form $LHS \rightarrow RHS$ then the noun/topic that is the RHS is added to the recommendation list. This process terminates when a set of k recommendations have been generated.

3.4 Recommendation Feedback and Case Retention

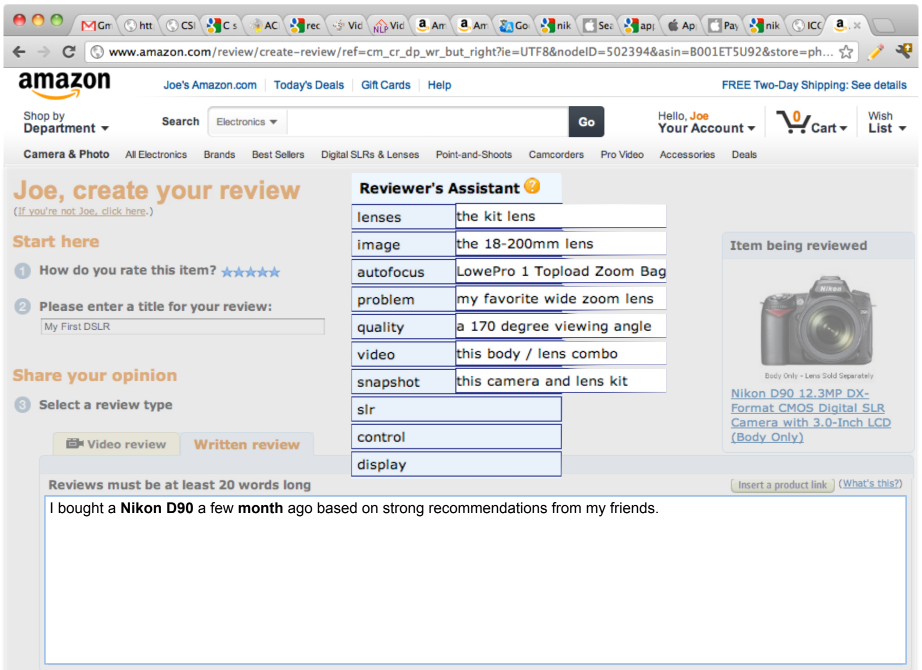
Each recommendation cycle refreshes the current set of suggestions for the user. Each of these recommendations can also be expanded to reveal the review fragments that formed the basis of the rule that led to the recommendation, thereby providing additional context for the user; for example, a recommendation for the topic *resolution* might show a review fragment such as *"... camera boasts an impressive 12 mega-pixel resolution..."*. By selecting a recommendation the user can directly add it to their review as a starting point for their own opinion.

There are a number of opportunities for learning to occur in the current system. Obviously, each completed review can be retained as a new case in the review case base. But in addition, it may also be possible to adapt the recommendation process by learning from direct user feedback; for example, as users select/ignore recommendations this can be used to reinforce/weaken future association rule patterns; this is left as a matter for future work.

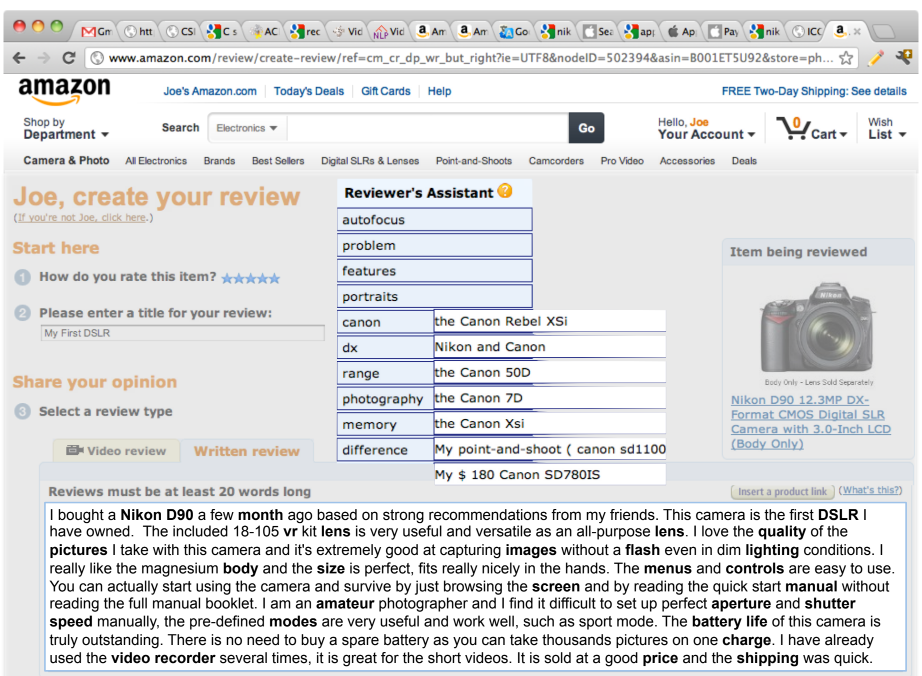
3.5 The Reviewer's Assistant in Action

Figure 3 shows the Reviewer's Assistant in action for our user, Joe, who is reviewing a recently purchased Nikon D90 SLR camera on Amazon. Joe is presented with the usual Amazon review creation screen and the figure shows the Reviewer's Assistant overlay; the RA widget can be dragged to any suitable location on screen. The RA presents a dynamic set of updating review suggestions (in this case we show the topic-based version of the RA). Figure 3a shows some of the suggestions presented to Joe during the early stages of the review. In this case we see a number of suggestions for some common review topics for this product, including the *lens*, the *image* and *video* capability. As shown, Joe can view review fragments that relate to a particular topic by mousing-over the topic. For example, in this case the fragments *"the lens kit"*, *"the 18-200mm lens"* etc. are displayed for topic *"lens"*.

In Figure 3b we see a snapshot towards the end of the review writing. This time Joe is presented with additional topics, many of which are more specialised



(a) Suggestions made at the beginning of the review writing process.



(b) Suggestions made toward the end of the review.

Fig. 3: The Reviewer's Assistant in action on Amazon.

or not uniquely related to the specific product to provide the reviewer with an opportunity to broaden their review. Note also that as the reviewer writes their review, sentences that cover recommended topics are highlighted by emphasising the topic terms in the sentences. In Figure 3b we can see that the user's review covers a range of topics that have been suggested, including *lens, aperture, shutter, battery life,* etc.

4 Live-User Evaluation

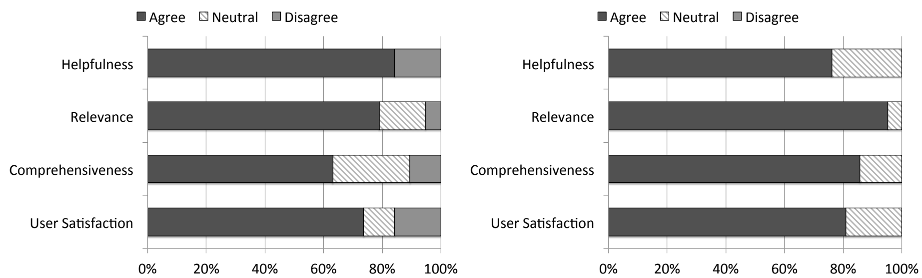
The Reviewer's Assistant, as described above, is designed to support users as they create user-generated reviews on sites like Amazon, TripAdvisor, and iTunes, by suggesting hints to users about the type of product features that they may wish to include in their reviews. We have presented two variations with respect to the type of recommendations made, one based on actual review fragments (*non-topic*) and one in which these fragments are first mapped to a set of well-defined topics, which are then recommended (*topic*). The following evaluation has three separate parts and in each, we pay particular attention to performance differences between the *non-topic* and *topic* variations of the RA, if any. In the first part, we describe the results of a live-user study focusing on how participants used the RA plugin and their feedback with respect to the utility of the recommendations and their overall satisfaction with the experience. In the second part of the analysis, we perform an objective analysis of the resulting reviews considering the depth and breadth of coverage offered by these reviews with respect to important product features. Finally, in part three, we perform a comparison of a subset of the above reviews and a set of comparable Amazon reviews, using a set of "expert" users to rate the helpfulness of these different reviews in order to better understand if the use of the RA leads to any improvements in review quality.

4.1 Usage Analysis

For the first part of this experiment we recruited 40 test users, 26 male and 14 female. 11 of the 40 participants had written at least one online review in the past and the majority had purchased products online through stores like Amazon and iTunes. We restricted our target product domain to that of digital cameras on Amazon and configured the RA plugins (*non-topic* and *topic*) accordingly; we chose this product domain because all users had at least some experience with this type of product.

The participants were randomly divided into *non-topic* and *topic* groups; exactly 21 participants (52.5%) had access to the *non-topic* version of the RA whilst the remaining 19 (47.5%) used the *topic* version. Each user was asked to select a product of interest and to write a review for this product; they were provided with a brief initial tutorial on the RA, the purpose of its suggestions, and how they might avail of them if appropriate.

During the trial user actions were logged as they completed their reviews and availed of the RA suggestions. At the end of the trial each user completed a short post-trial questionnaire in order to rate the RA under four key areas: 1) helpfulness – were the RA suggestions generally helpful? 2) relevance – were the suggestions relevant in the context of the review being written? 3) comprehensiveness – did the suggestions broadly cover the product being reviewed? 4) overall satisfaction – was the participant satisfied with the overall experience provided by the RA?



(a) Topic suggestions. (b) Non-topic suggestions

Fig. 4: User Feedback.

The results of this questionnaire are shown in Figure 4 and are largely positive with respect to both the *topic* and and *non-topic* variations. For example, we can see that overall about 75% – 85% of users found the RA to be helpful, with the higher percentage pertaining to the *topic* variation. Interestingly, this advantage for *topic* is reversed when we look how relevant users found the recommendations to be. In this case we can see that, while 90% of the *non-topic* users rated the recommendations to be relevant, only 80% of the *topic* users rated their recommendations as relevant. Similarly, the feedback in terms of recommendation comprehensiveness also favours the *non-topic* variation, with scores of 85% versus 65% for *non-topic* and *topic* users, respectively.

The reason for this seems to be related to differences in how recommendations are managed after a user has covered a particular feature in their review. For example, in the case of the *non-topic* version, once a user writes about a specific feature, say *auto-focus*, this exact feature will be removed from the recommendation list and will not be suggested again in the future, but related features such as *zoom* or *focal-length* can be suggested. The same is true for the *topic* variation of RA, except that by removing the general topic, in this case *lens*, instead of the exact term *auto-focus*, lens-related features will not be suggested any more and *lens* will be replaced by another topic instead. In retrospect it appears that this particular recommendation filtering approach may have been overly restrictive and we will consider alternatives as a matter of future work.

Finally, in relation to the post-trial questionnaire we can see that overall there is strong user-support for the RA. Between 75% (*topic*) and 80% (*non-topic*) of users indicated that they were satisfied overall with the system.

4.2 Topic Coverage

We now consider the type of reviews that are produced. For example, is there any evidence that the *topic* and *non-topic* variations lead to quantitative differences in review quality? In this part of the evaluation we consider review quality in terms of the breadth and depth of topical coverage. In other words we can evaluate reviews based on the number of unique topics that they contain (*breadth*) and the average length of sentences on a given topic (*depth*). We can also measure the *redundancy* of a review as the average length of sentences that do not refer to well-defined topics.

More formally, the *breadth* of a review r with respect to topic set T is defined as the number of topics covered by that review, see Equation 1.

$$Breadth(r, T) = \mid \{t \in T \mid \exists s \in r : Cover(s, t)\} \mid \tag{1}$$

$Cover(s, t)$ is *true*, if topic t *is covered by* sentence s. A sentence *covers* a topic, if at least one synonym for (or member of) the topic is contained in it. In order to give proper semantics to the mathematical notation of $\in$ we assume that sentences are represented by a collection of all words, and reviews are represented by a collection of sentences. Note, one sentence might cover more than one topic.

The *Depth* of a review r with respect to topic set T is the average number of words to describe each topic covered by the review r, see Equation 2.

$$Depth(r, T) = \frac{\sum_{\{s \in r \mid \exists t \in T : Cover(s,t)\}} Length(s)}{Breadth(r, T)} \tag{2}$$

where the number of words in a sentence s is denoted by $Length(s)$.

Finally, the *Redundancy* of a review r with respect to topic set T is defined as the total length of sentences that do not cover any topic.

$$Redundancy(r, T) = \sum_{\{s \in r \mid \neg \exists t \in T : Cover(s,t)\}} Length(s) \tag{3}$$

	topic	non-topic
Average Breadth*	10.42	7.62
Average Depth	10.69	10.53
Average Redundancy	9.68	10.24
Average Length	113.58	90.43

Table 1: Breadth, Depth, and Redundancy; * indicates significance at 0.05.

Thus we analyse the review texts of the 40 reviews produced during the above trial and compute their breadth, depth, and redundancy characteristics with reference to the defined set of product topics used for the digital camera domain. The results are presented in Table 1. We can see that while both techniques perform similarly in terms of review depth (10.69 for *topic* versus 10.53 for *non-topic*), the reviews produced with the *topic* version of RA tend to offer significantly broader coverage (a breadth of 10.42 for *topic* versus only 7.62 for *non-topic*) with less redundancy (9.68 for *topic* versus 10.24 for *non-topic*).

Of course this approach provides only a superficial analysis of review quality and it is not clear whether these depth, breadth and redundancy characteristics have any significant bearing on the ultimate perception of review quality or helpfulness. Thus, we analysis performance along these dimensions in the next section.

4.3 Review Quality

Ultimately the best test of the RA approach is to consider the quality of the resulting reviews in order to understand whether users find them to be helpful, for example. Even better is if we can compare our test reviews to a benchmark in terms of quality. This is the aim of this final evaluation section.

We collected 2 sets of reviews with similar lengths. The first set were chosen at random from the reviews written by participants of the RA trial above. We collected 10 random reviews written using the help of the RA with *topic* and another 10 written using the help of the RA with *non-topic*. For our second set, we selected two groups of Amazon reviews to serve as a benchmark, against which to judge the quality of the RA reviews. One group was chosen at random from among the most helpful Amazon camera reviews. We picked 10 reviews (*Amazon+*) that had a helpfulness score of at least 0.7 (meaning 70% of raters considered them helpful); in fact, the average helpfulness score for these reviews was 0.9 and thus we can view these as examples of very high quality product reviews written without the aid of RA. Next we chose another group of 10 random Amazon reviews (*Amazon-*), but this time we picked reviews that had a helpfulness score of less than 0.7; the average helpfulness score for reviews in this group was 0.41 and thus represent examples of lower quality reviews written without the help of RA.

Next, we recruited 15 reviewer "experts" (with a good understanding of the digital camera space) and asked them to perform a blind review of a random sample of reviews from the four sets above (*topic*, *non-topic*, *Amazon+*, and *Amazon-*). In each case we asked the experts to rate the reviews on a 5-point scale in terms of 1) helpfulness – how helpful did they think the review would be to others? 2) completeness – did the review provide a reasonably complete account of the product in question? 3) readability – was the review well written and readable? In total each test review was reviewed by 3 different experts. Finally, we calculated the average helpfulness, completeness, and readability ratings across each of the 4 review groups and also calculated their average breadth, depth and redundancy scores based on the approach taken previously.

	topic	non-topic	Amazon+	Amazon-
Helpfulness*	3.90	3.90	3.33	3.07
Completeness*	3.67	3.57	2.67	2.53
Readability	3.60	3.60	3.80	3.33
Average Breadth**	11.30	8.60	5.90	7.20
Average Depth	11.90	11.21	15.57	12.49
Average Redundancy**	5.50	14.30	22.40	21.00

Table 2: User Evaluation; * indicates significant difference between topic/ non-topic and Amazon+/ Amazon-; ** indicates significant difference between topic only and Amazon+/ Amazon-.

The results are presented in Table 2 and show a significant positive benefit accruing to the RA in a number of important respects. For instance the average helpfulness rating of RA reviews (3.90 for both RA versions) is greater than the helpfulness rating for *Amazon+* (3.33) and *Amazon-* (3.07). Similarly, we can see clear benefits for the RA variations in terms of review completeness (3.67 and 3.57), when compared to *Amazon+* (2.67) and *Amazon-* (2.53). Both of these helpfulness and completeness benefits (RA versus Amazon) are statistically significant at the 0.05 level; statistically significant differences were not found in terms of review readability.

Table 2 also shows how the *topic* version of the RA leads to reviews that have a greater topical breadth and reduced redundancy compared to *non-topic*, *Amazon+*, and *Amazon-*; once again these differences are statistically significant at the 0.05 level. For example, on average, the *topic* reviews cover 11.3 topics per review compared to only 5.9 and 7.3 topics per review for *Amazon+* and *Amazon-*. In addition, on average, the *topic* reviews contain very little redundancy (5.50 words per review) compared to much higher redundancy rates for *Amazon+* (22.4) and *Amazon-* (21) and even *non-topic* (14.3). It is worth noting that, though we did not find statistically significant differences, both groups of Amazon reviews seem to enjoy improved topical depth when compared to the RA groups.

4.4 Discussion

The above results tell an interesting and compelling story. While there may be little to choose between the *topic* and *non-topic* variations of the RA at present, it is clear that users found the RA system to be helpful and useful when it comes to supporting the review writing process, Moreover, the resulting reviews were rated more highly that their Amazon counterparts in a blind study of review quality.

As always there are limitations to be considered when evaluating the significance of these results. For a start the review domain was limited to digital cameras and the user studies were limited to focus groups of 40 users. Nevertheless the key results were found to be statistically significant rather than chance occurrences. Moreover, we have found no reason to suspect that by focusing on digital cameras we have in any way biased or skewed the evaluation. Certainly digital cameras are a popular class of online product sales and attract a critical mass of user reviews. Moreover they share similar characteristics (feature-based descriptions, feature tradeoffs, leading to complex purchasing decisions) with other classes of products and services (travel, gadgets, etc.). Given the above we feel confident that these results bode well for the promise of the RA and as such provide a clear demonstration of the value of a case-based approach to harness online experiences.

5 Conclusions

This work was inspired by early work on the GhostWriter systems [3, 9], which highlighted the potential for web experiences and case-based reasoning to support users when creating user-generated content, whether in the form of adverts or product reviews. The main contribution of the work presented in this paper is twofold. Firstly, we have extended the original Ghostwriter approach by incorporating a combination of association rule mining and topic extraction to generate review recommendations that are more likely to match key product features as important review targets. Secondly, we have presented a comprehensive evaluation of the RA system, focusing on the overall user experience and a benchmarked study of real review quality. The results show that users find the RA to be useful and the resulting reviews are rated more highly than comparable Amazon reviews, even when compared against a set of best quality reviews.

In terms of future work there are a number of important possibilities. First of all, our current approach to topic extraction is very simply, modeling topics based on simple synonym sets, for example. Our next steps include exploring the use of automatic topic-detection and extraction techniques which will allow for a more sophisticated topic modeling approach. In addition, the RA currently does not consider the sentiment of review fragments during recommendation; for example, a given reviewer might speak positively about a particular product feature whilst another reviewer may speak negatively. This type of information can be useful when selecting recommendations, for example, by guiding users to review more controversial features of the product. Finally, the RA currently focuses on past review experiences that match the target product being reviewed. This potentially limits the scope of experiences that can influence recommendations and it is worth considering whether drawing on reviews from similar products is likely to be of benefit. For example, when reviewing a specific compact camera by Nikon it might be worth focusing on other compacts by Nikon or other manufacturers. We will consider this in future work by relaxing the similarity metric that is used during review retrieval.

6 Acknowledgments

This work is supported by Science Foundation Ireland under grant 07/CE/I1147.

References

1. R. Agrawal, T. Imieliński, and A. Swami. Mining Association Rules between Sets of Items in Large Databases. *ACM SIGMOD Record*, 22(May):207–216, 1993.
2. D. W. Aha, L. A. Breslow, and H. Muñoz-Avila. Conversational Case-Based Reasoning. *Applied Intelligence*, 14(1):9–32, 2001.
3. D. Bridge and P. Healy. The GhostWriter-2.0 Case-Based Reasoning system for making content suggestions to the authors of product reviews. *Knowledge-Based Systems*, 29(0):93 – 103, 2012.
4. D. Bridge and A. Waugh. Using experience on the read/write web: The ghostwriter system. In D. Bridge, E. Plaza, and N. Wiratunga, editors, *Procs. of WebCBR: The Workshop on Reasoning from Experiences on the Web (Workshop Programme of the 8th International Conference on Case-Based Reasoning)*, pages 15–24, 2009.
5. J. A. Chevalier and D. Dina Mayzlin. The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43(3):345–354, 2006.
6. C. Dellarocas, M. Zhang, and N. F. Awad. Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing*, 21(4):23–45, November 2007.
7. V. Dhar and E. A. Chang. Does chatter matter? the impact of user-generated content on music sales. *Journal of Interactive Marketing*, 23(4):300–307, 2009.
8. R. Dong, K. McCarthy, M. P. O'Mahony, M. Schaal, and B. Smyth. Towards an intelligent reviewer's assistant: Recommending topics to help users to write better product reviews. In *Procs. of IUI: 17th International Conference on Intelligent User Interfaces, Lisbon, Portugal, February 14-17, 2012*, pages 159–168, 2012.
9. P. Healy and D. Bridge. The GhostWriter-2.0 System: Creating a Virtuous Circle in Web 2.0 Product Reviewing. In D. Bridge, S. J. Delany, E. Plaza, B. Smyth, and N. Wiratunga, editors, *Procs. of WebCBR: The Workshop on Reasoning from Experiences on the Web (Workshop Programme of the Eighteenth International Conference on Case-Based Reasoning)*, pages 121–130, 2010.
10. N. Hu, L. Liu, and J. Zhang. Do online reviews affect product sales? the role of reviewer characteristics and temporal effects. *Information Technology and Management*, 9:201–214, 2008. 10.1007/s10799-008-0041-2.
11. S.-M. Kim, P. Pantel, T. Chklovski, and M. Pennacchiotti. Automatically assessing review helpfulness. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2006)*, pages 423–430, Sydney, Australia, July 22-23 2006.
12. Y. Liu, X. Huang, A. An, and X. Yu. Modeling and predicting the helpfulness of online reviews. In *Proceedings of the 2008 Eighth IEEE International Conference on Data Mining (ICDM 2008)*, pages 443–452, Pisa, Italy, December 15-19 2008.
13. M. P. O'Mahony and B. Smyth. Learning to recommend helpful hotel reviews. In *Proceedings of the 3rd ACM Conference on Recommender Systems (RecSys 2009)*, pages 305–308, New York, NY, USA, October 22-25 2009.
14. G. Wu, D. Greene, and P. Cunningham. Merging Multiple Criteria to Identify Suspicious Reviews. In *Proc. 4th ACM Conference on Recommender Systems (RecSys'10)*, pages 241–244, 2010.
15. G. Wu, D. Greene, B. Smyth, and P. Cunningham. Distortion as a Validation Criterion in the Identification of Suspicious Reviews. In *Proceedings of the First Workshop on Social Media Analytics*, SOMA '10, pages 10–13, New York, 2010.
16. F. Zhu and X. M. Zhang. Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of Marketing*, 74(2):133–148, 2010.