Recommender System Based on Consumer Product Reviews

Silvana Aciar
University of Girona Campus Montilivi, Building P4, 17071 Girona, Spain
saciar@eia.udg.es

Debbie Zhang
University of Technology, Sydney, PO Box 123 Broadway 2007, Australia
debbiez@it.uts.edu.au

Simeon Simoff
University of Technology, Sydney, PO Box 123 Broadway 2007, Australia
simeon@it.uts.edu.au

John Debenham
University of Technology, Sydney, PO Box 123 Broadway 2007, Australia
debenham@it.uts.edu.au

Abstract

Consumer reviews, opinions and shared experiences in the use of a product is a powerful source of information about consumer preferences that can be used in recommender systems. Despite the importance and value of such information, there is no comprehensive mechanism that formalizes the opinions selection and retrieval process and the utilization of retrieved opinions due to the difficulty of extracting information from text data. In this paper, a new recommender system that is built on consumer product reviews is proposed. A prioritizing mechanism is developed for the system. The proposed approach is illustrated using the case study of a recommender system for digital cameras.

1. Introduction

Recommendation systems are programs which attempt to predict items that a user may be interested in, given some information about the user's profile. Most existing recommender systems use collaborative filtering methods, content-based methods or hybrid filtering methods that combine both two techniques. Collaborative filtering methods base recommendations on other users' preferences. By contrast, content-based methods use information about an item itself to make suggestions. Collaborative filtering approaches that produce recommendations by comparing a consumer's previous selections with other consumers who have made similar selections. Collaborative filtering systems overcome many shortcomings of contents based systems. These systems use a collection of historical rating data of m users on n products as input, which are collected by asking users to input the rating of the products as numerical values [1]. However, many consumers prefer to use free form of text to express their opinions. Produce review forums and discussion groups are popular ways for consumers to exchange their experiences with a product [2][3][4]. These consumer reviews and opinions published in various sources, including virtual community logs, discussion boards and e-commerce sites. There is growing evidence that such forums inform and influence consumers' purchase decisions [5, 6]. Despite the importance and value of such information, there is no comprehensive mechanism that formalizes the opinions selection and retrieval process and the utilization of retrieved opinions due to the difficulty of extracting information from text data. Adomavicius provided an overview of recent development of recommender systems [7]. According to his review, the recommender systems that utilize review comments using text mining techniques are yet to be developed. Ricci [8, 9] proposed to utilize review comments for product description and user behavior study. He believed the review comments could be widely used in recommender systems and result in better recommendations.

In this paper, a recommender system that utilizes online consumer opinions about the products is presented. The review comments could come from chat rooms or online discussion forums. Text mining techniques are employed to extract useful information from review comments. Ontology has also been defined to translate the review information into a form suitable for utilization by the recommender system. A ranking mechanism for prioritizing that information with respect to the consumer level of expertise in using that product has been developed.

Figure 1 shows the system structure of the proposed recommender system. Different from other recommender systems, the system uses consumer review comments that are in free form text as input. In order to demonstrate the effectiveness of this approach, a recommender system for a specific type of products - digital cameras has been developed. The rest of the paper is organized as follows. The next section describes the potential and value of consumers'

reviews as information source in a recommender system and the measures used to calculate a rating of a product based on such information. Section 3 describes a case study in the digital camera domain and finally Section 4 concludes the paper and provides directions for future research.
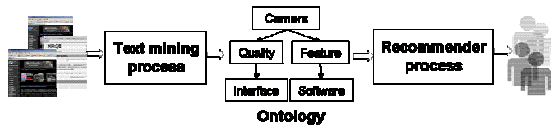

Figure 1. System Structure of the Recommender System that Utilizes Consumer Opinions as the Input Information Source

2. Prioritizing system for consumer reviews

These reviews provide valuable source of information for recommender systems. However automating the acquisition of such information requires innovative technological solutions. The consumer reviews are textual and unstructured sources that are particularly difficult to acquire. The effective selection/retrieval of consumer opinions entails several tasks, such as:

1) Representation of the information in a common format (ontology generation).
2) Computing the rating of the product from the opinions.
3) Selecting the most relevant opinion and making recommendations in response to a user request.

2.1. Representation of consumer reviews

A typical review comment could be like:
Canon PowerShot A530 is a very good camera with very good image quality. 5 MP is enough for very sharp pictures in almost every condition. Only negative is plastic body but considering the prize it is by far one of the most valuable cameras to buy. 170€ for 4x zoom, 5MP and very good handling! Many useful features for great pictures... After trying different higher prize cameras I was impressed by the speed of the Af and the typical Conon menu and functions. This is not the "smallest ever 3x zoom 8 MP Camera" but very good thing to work with. ISO 400 and 800 does not look really good.

The goal of this step is to find a suitable tool for extracting the information contained in the text and converting it into structured data, such as a form depicted in Figure 2. Identifying an appropriate representation of consumer opinions that can be used in the system is a key problem.
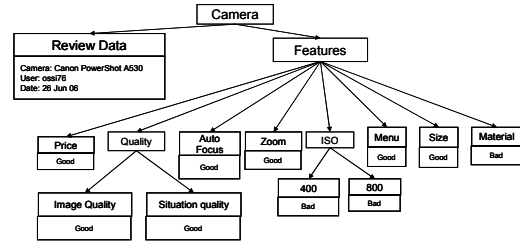

Figure 2. A Representation of the Information in the Review of Canon PowerShot A530

One way to convert these opinions to a structured form is to use translation ontology, which is typically used as a form of knowledge representation and sharing. An ontology is a collection of concepts and their relationships that can collectively provide an abstract view of an application domain [10]. Review comments are firstly mapped into ontologies to allow the ranking calculations become possible. In this application, the ontology contains two main parts: Opinion Quality and Product Quality, which summarize the consumer skill level and the consumer experience with the product in the review, respectively. Figure 3 shows the general structure of the ontology.
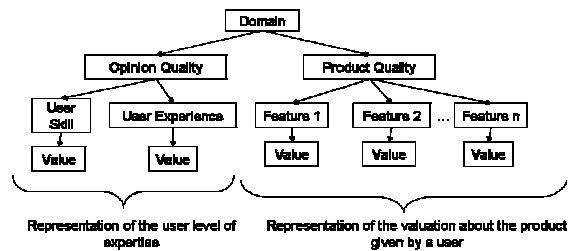

Figure 3. Structure of the Ontology used in the Recommendation from Consumer Opinions Applications

The Opinion Quality includes several variables to measure the opinion provider's expertise in the product. The Product Quality represents the opinion provider's valuation of the product features, which is highly domain specific. Section 3 presents the ontology developed for digital cameras reviews that are used in the study cases of this paper.

2.2. Rating the consumer skill level

The review comments were given by people with diverse experience and skill levels. In general, people who have longer history of using the product can provide more professional opinions. Therefore, these diverse opinions should not be treated equally. The opinions from more experienced people should be

taken in account to a greater extent than those from people with little knowledge of the product. Opinion Quality (OQ) is defined to evaluate the weighting value of opinions according to the opinion providers' expertise.

Definition 1. Opinion Quality (OQ) is the sum of the weight w_j, given for each variable j representing the skills and experiences of consumer i divided by the number of variables representing the information about consumer's skill and expertise provided in the ontology.

$$OQ_i = \frac{\sum_{j}^{n} w_j}{n} \qquad (1)$$

The Opinion Quality is calculated by the values stored in the corresponding part of the ontology. Detail definition of the variables is given in section 3.2. An Opinion Quality value is calculated for each piece of comment.

2.3. Product quality ranking

The product is ranked according to the consumer comments for each feature. Due to the difficulties of quantification of user valuation from texture data, each feature from the comment can only be assigned either "Good" or "Bad", which is calculated as "1" or "-1" respectively. For each feature, a Feature Quality is calculated, which is a function of consumer valuation and Opinion Quality.

Definition 2. Feature Quality (FQ): The quality value for each feature of the product in a review is the rating multiply by the Opinion Quality value of the consumer

$$FQ_f = r * OQ_i \qquad (2)$$

2.4. Selection of the relevant opinion and making recommendations in response to a user request

When a user requests the evaluation of a particular product based on certain features, the Overall Feature Quality is calculated from the reviews that contain the valuation of this feature.

Definition 3. Overall Feature Quality (OFQ) is the global valuation of the feature from all reviews, which is calculated by the average value of Feature Quality.

$$OFQ_f = \frac{\sum (Scalingfactor * FQ)}{NumberOfOpinions} \qquad (3)$$

Here Scaling Factor is used to do the minor adjustment of the user valuation, which can be set to:

$$Scalingfactor = \frac{1}{n} \qquad (4)$$

n is the number of all the features rated by the consumer. Each review rated different number of features so n could be different. To provide the user with a comprehensive valuation of the product quality in related to the requested features, an Overall Assessment score is defined.

Definition 4. Overall Assessment (OA) provides a final score of the product based on the valuation of each feature. It is calculated as the sum of all OFQ (calculated by equation (3)) multiplied by the Importance Index.

$$OA = \sum OFQ * Importance\ Index \qquad (5)$$

The Importance Index measures the different influence of the features to consumer's decision making, which can be assigned in two ways: according to the importance of the feature expressed in the user request or by the frequency that the features have been rated in the consumer reviews.

3. Case study

Case study was conducted using digital cameras. Data from the Digital Photography Review (www.dpreview.com) were chosen where each day consumers visit this page to rate and add opinions about different digital cameras. In this section, detail calculations of a user request were shown and recommendation was given based on the calculations.

3.1. Representation of the consumer reviews - Digital camera ontology

In computing ontology is a specification of an abstract, simplified view of the world that is wised to represent for some purpose [10]. Therefore, ontology defines a set of representational terms called concepts. Interrelationships among these concepts describe a target world. In this research, an ontology has been developed for digital camera domain. Each concept in the ontology was obtained analyzing the reviews from the consumers of different digital cameras from www.dpreview.com. Consumers can choose any digital camera and rate it on a scale of half start to four starts. They can also write free form text reviews about the camera. For the construction of digital camera reviews ontology, first was made a list of all possible objects necessary to cover given cameras reviews. This

possible list should include different digital cameras such as Canon, Sony, etc. Furthermore, different cameras can be qualified by features such as size, zoom, lens, quality picture, etc. This information is represented by the concept "Features". And the different consumer's reviews can be qualified by opinions from beginners, professionals and by the level of expertise using digital cameras. This information is represented in the ontology by the concept "Opinion Quality".

3.2. Obtaining the Opinion Quality (OQ) and Feature Quality (FQ)

The Opinion Quality is calculated by equation (1) in section 2.2. Table 1 presents the weighting value of each variable defined in the equation. The OQ values for each consumer can be calculated using Table 1. Customer John's OQ value is calculated as:

$$OQ_{jhon} = \frac{0.5 + 0.7 + 0.7 + 0.5}{4} = 0.6$$

Feature Quality (FQ) value for each feature rated by the consumers is also calculated. For example as shown in Table 2, John gave the value "good" or "bad" for each feature of the digital camera Sonny361 and his OQ value is 0.6. As described in previous sections, by assigning the value 1 for "good" and -1 for "bad" in equation 2, the Feature Quality for each feature in John's opinion are calculated. The same process has been applied to all consumers. The OQ and FQ for each review comment are calculated off-line to achieve quick response to the user requests. The recommender system requires from the user to input the model of the camera he (she) is interested and selects the features that he (she) is most concern. The features in the selection panel are the same set of features that is covered by the ontology. For example, a user request **"I would like to know if Sony361 is a good camera, specifically its interface and battery consumption"** is presented. Three keywords (Sony361, interface and battery) can be identified. Firstly, only the opinions for Sony361 are selected. In this case study, there are three opinions about Sony361's cameras: John's opinion, Karen's opinion and James's opinion.

Then the OFQ of each feature is calculated using equation (3).

$$OFQ_{interface} = \frac{1}{2}\left(\frac{1}{4}*(-0.6) + \frac{1}{4}*(-0.75)\right) = -0.165$$

$$OFQ_{battery} = \frac{1}{4}*0.75 = 0.18$$

Table 1. Variables representing the consumer level expertise in using a digital camera

	Value	weight (wi)
Consumer Skill	Beginner	0.5
	Advanced	0.7
	Professional	0.9
Consumer Experience		
Time to use this camera	Day	0.3
	Week	0.5
	Month	0.7
	Year	0.9
Time to use Digital Camera	Day	0.3
	Week	0.5
	Month	0.7
	Year	0.9
Number of different Cameras	One	0.3
	Two	0.5
	Three	0.7
	(+) Three	0.9

Table 2. Information about John's opinion

Consumer:	John
Camera:	Sony361
OQjhon:	0.6
Size:	good
Interface:	bad
Documentation:	good
Zoom:	good

The Overall Assessment for the digital camera Sony361 based on the two features requested is obtained using equation (5). The Importance Index was calculated in two ways. For the case of using the importance index from the user request where the user has expressed that the interface is more important than the battery, so the value of 1 is assigned for interface and 0.5 for battery. Using these values the OA for Sony361 camera is:

$$OA = -0.165*1 + 0.18*0.5 = -0.075$$

In the case of no user preference is given, the importance index are calculated based on the frequency of the feature being reviewed:

$$Importance\ Index = \frac{n}{N} \qquad (6)$$

Where n is the number of time that the feature appears in the reviews and is the total number of reviews. Using equation (6), the OA for Sony361 camera is:

$$Importance\ Index\ Int\,erface = \frac{6}{10} = 0.6$$

$$Importance\ Index\ Bat\,tery = \frac{2}{10} = 0.2$$

$$OA = -0.165*0.6 + 0.18*0.2 = -0.063$$

Assigned the value "Good" for OA and OFQ > 0 and "Bad" for OA and OFQ< 0 the Sony361 camera is "Bad" according to consumers' opinions. The response for the user request is shown in Figure 4.

The best camera with the features the user concern is also recommended. The same process is applied to all other cameras review. CannonTW45 is recommended considering this information. The complete recommendation is show in Figure 5.
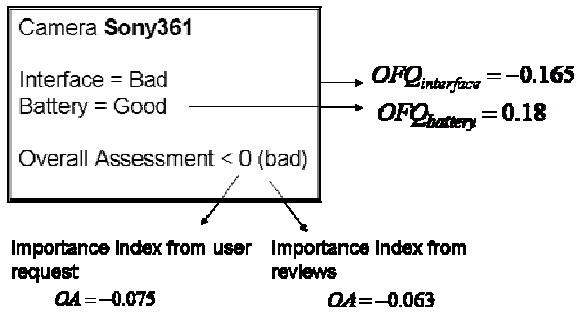

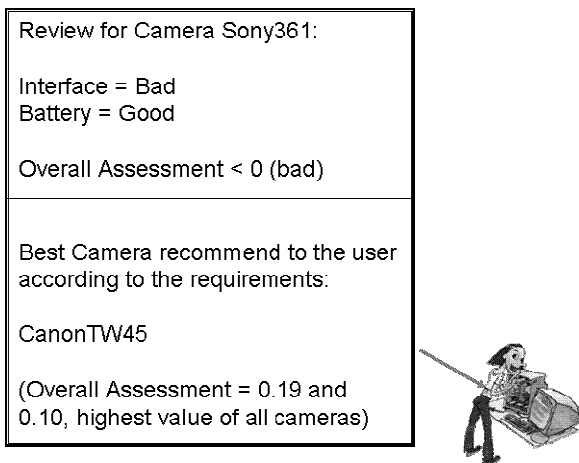**Figure 4. Recommendation in Response for a User Request from the Consumers' Opinions**

Figure 5. Final Recommendation generated from the consumer's opinions about Digital Cameras

4. Conclusion

This paper proposed a novel approach to creating recommendations in recommender systems, which utilizes online consumer review comments. To the best of our knowledge, this is the first attempt to build a recommender system based on review comments in free form text. A ranking mechanism for prioritizing the product quality with respect to the consumer level of expertise and the rating given to some features of the product has been developed. The approach uses domain ontology to translate the information into a form that is suitable for processing by the recommender system. Such ontology has been defined for the domain of digital camera reviews and has been used for demonstration of the work with some examples. A set of measures such as Opinion Quality (OQ), Feature Quality (FQ), Overall Feature Quality (OFQ) and Overall Assessment (OA) have been defined to select the relevant reviews and provide the best recommendation in response to a user request. The recommendation is given based on these measurements. In the case study presented in this paper, the mapping of review comments into ontologies was conducted manually. The future development of the approach considers the automation of, this mapping process by using text mining technique. Also, the implemented system should be evaluated with the intended consumer groups.

5. References

[1] W. Yang, Z. Wang, M. You, "An improved collaborative filtering method for recommendations' generation", in Proc. of IEEE International Conference on Systems, Man and Cybernetics, 2004.
[2] C. Dellarocas, "The digitization of Word-of-Mouth: Promise and challenges of online feedback mechanisms", Management Science, 49 (10): pp. 1407-1424. 2003.
[3] C. Dellarocas, "Strategic manipulation of Internet opinion forums: Implications for consumers and firms", Working Paper, Sloan School of Management, MIT, Cambridge. 2004
[4] N. Curien, E. Fauchart, G. Laffond and F. Moreau, "Online consumer communities: escaping the tragedy of the digital commons", in, Internet and Digital Economics, Brousseau E. and N. Curien (eds). Cambridge University Press, 2006.
[5] S. Senecal and J. Nantel, "The influence of online product recommendations on consumers' online choices", Journal of Retailing, 80, Elsevier, pp. 159-169.2004
[6] J. Chevalier and D. Mayzlin, "The effect of word of mouth on sales: Online book reviews", NBER Working Paper Series, National Bureau of Economic Research, USA. 2003.
[7] G. Adomavicius, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions", IEEE Transactions on Knowledge and Data Engineering, vol 17, no. 6, 2005.
[8] R. Wietsma, F. Ricci, "Product Reviews in Mobile Decision Aid Systems", Workshop on Pervasive Mobile Interaction Devices, in conjunction with Pervasive 2005, Munich, Germany, May 11, 2005.
[9] F. Ricci and R. T. A. Wietsma, "Product Reviews in Travel Decision Making", Information and Communication Technologies in Tourism Proceedings of the International Conference in Lausanne, Switzerland, Pages 296-307, Springer Verlag, 2006.
[10] T.R. Gruber, "A translation approach to portable ontology specifications", Knowledge Acquisition, 5(2):199–220. 1993.