Abstract

Product recommender systems aim to support consumers in making buying decisions. However, such a support requires considering the consumer behaviour in making buying decisions. In this paper, we deduce design requirements for utility-based recommender systems from the theory of consumer information behaviour and present empirically findings from experiments conducted with a prototypical implementation of the proposed requirements. The empirical examination shows that our recommender system has a high predictive validity.

Keywords: Utility-based Recommender System, Consumer Information Behaviour, Design Requirements, Laboratory Experiment

1 Introduction

One of the most important advances of electronic commerce is the reduction of search costs invested to find an adequate product (Bakos 1997; Hinz and Eckert 2010). Though, the reduction of search costs requires the support of recommender systems, due to the high amount of product and producer information available on the Internet. Recommender systems have been widely examined in recent years. Burke for instance divides existing recommender systems based on their input data and their technique into five categories: collaborative filtering, content-based filtering, demographic filtering, knowledge-based filtering and utility-based filtering (Burke 2002).

In most research prototypes as well as practical implementations (e.g. Amazon.com), collaborative filtering or content-based filtering are applied as recommendation technique (Herlocker et al. 2000; Xu et al. 2005; Mehta et al. 2007). Since these methods are afflicted with some drawbacks like the cold-start problem (Burke 2002), alternative methods like knowledge-based filtering and utility-based filtering have been focussed in recent research (DeBruyn et al. 2008; Scholz 2008; Wilson et al. 2009).
Utility-based recommender systems try to elicit current consumer preferences to predict appropriate recommendations. Recommender systems based on collaborative or content-based filtering apply historical data to estimate consumer preferences. Since historical data might become obsolete, utility-based recommender systems are suitable especially for recommending products sparsely purchased in a consumer’s life, such as notebooks or cars. The core of each utility-based recommender system is a method for measuring consumer preferences. Several methods such as the self-explication approach (Cao and Li 2007; Theetranont et al. 2007), conjoint analysis (Scholz 2008), and neural networks (Schneider 2005) have already been implemented for measuring preferences in a recommender system. Depending on the measurement method (technique in the sense of Burke), user have to directly or indirectly specify their preferences for a set of attributes which characterise the product type (e.g. notebooks) the user is looking for (input date in the sense of Burke). The self-explication approach for example requires a direct specification of attribute preferences. In contrast, a conjoint analysis requires the specification of product preferences which are used to conjointly assess attribute preferences. However, there is no theoretical foundation in any of these papers of how to present information required in the process of measuring preferences.

In this paper, we focus on findings from the theory of consumer information behaviour to deduce a list of design requirements for utility-based recommender systems. We furthermore present some empirically findings from a prototype in which all design requirements are implemented.

This paper is organised as follows. In the next section, we explain the research methodology for this study. Theoretical findings of consumer information behaviour are presented in section 3. Design requirements are deduced from these findings in section 4.1. The implementation of these requirements is briefly described in section 4.2. An empirical examination of the prototype in order to validate the requirements is presented in section 5. We conclude the paper in section 6 with limitations and implications for research and practice.

2 Methodology

The goal of this article is defined as the extraction of design requirements for utility-based recommender systems from the theory of consumer information behaviour. The starting point of our work is a theory which has been examined especially in the 1970s and 1980s. In order to test whether the design requirements are valid, an implementation of a prototype is imperative. The evaluation of the requirements and thus also the theory of consumer information behaviour is here conducted with a laboratory experiment. Our methodology is hence in line with design research as described by Nunamaker and Chen (1991).

We present some findings of the theory of consumer information behaviour within the next section. These findings are hypotheses which have been already falsified in recent years. Each hypothesis consists of two constructs, an antecedence $P$ and a consequence $Q$. The antecedence $P$ is reformulated as design requirement if and only if $Q$ is a consequence that supports the goal of the design artefact (Gehlert et al. 2009). The overall goal of a utility-based recommender system is to accurately predict recommendable products. This goal requires validly measuring consumer preferences. We hence can deduce design requirements from those hypotheses having a consequence
that supports a valid preference elicitation. Though, not all of the antecedences of these hypotheses are implementable in software\(^1\).

3 Consumer Information Behaviour

Consumers aim to make optimal buying decisions. Their behaviour in making buying decisions differs from product to product, though. Especially the intensity of searching information about alternatives is variable. However, a model (see Figure 1) which is adequate for each buying decision has been presented by Kotler and Keller (2008).

![Figure 1: Buying decision process (Kotler, Keller 2008)](image)

Recommender systems focus on information search and evaluation of alternatives. Utility-based recommender systems estimate the utility for each product based on elicited consumer preferences. Hence, these systems evaluate alternatives according to the utility they provide to a particular consumer. As mentioned above, methods for measuring preferences have been already investigated in several studies. To reliably measure consumer preferences these methods must be adapted according to the information behaviour of consumers. All of the preference measurement methods assume that the utility of a particular product is a function of the preferences a consumer has for attributes. Measuring preferences of as many attributes as consumers typically consider when making a buying decision is hence imperative to reliably measure preferences. The number of attributes considered in buying processes is an example of findings from research in consumer information behaviour.

Consumer information behaviour is here defined as the art and wisdom of gathering, storing and digesting information (Silberer 1981). This definition encompasses all cognitive processes influencing the information behaviour of consumers. Storage and digestion of information mainly depends on the process of gathering information whereas the collection itself is determined by several determinants. Information is here defined as data chunk consisting of the level of a particular attribute and a particular product.

Some listings of determinants are presented in marketing literature (Bettman 1978; Silberer 1981; Loudon and Della Bitta 1984; Wilson 2000). A theoretical based classification of information gathering determinants has been presented by Kuss (1987) and is used in this paper. Gathering information is determined by three main aspects:

- Person
- Task
- Situation

\(^1\) For example, if the antecedence is related to human attitudes, an implementation in software is mostly not possible.
The relationship between “person” and “task” defines the determinant “problem orientation”, whereas the relationship between “task” and “situation” determines the determinant “information offer”. The whole framework of information gathering determinants is shown in Figure 3.

![Figure 3: Determinants of information gathering (Kuss 1987)](image)

It is worth noting, that the framework is not complete at all, but appropriate to deduce design requirements for utility-based recommender systems. We briefly discuss each determinant of Figure 3 within the following subsections. Since information storage and digestion are determined by the behaviour of gathering information, these aspects do not provide any further information for design requirements and they are hence not discussed here.

### 3.1 Person

The consumer as person is characterised by socio-demographic as well as behavioural variables. In recent studies socio-demographic variables were mostly on focus, since they are much easier to collect than behavioural variables. Significant influences on the process of information gathering have been identified for age (Capon and Kuhn 1980; Roedder John 1999), income, education and profession (Fritz and Hefner 1981), and gender (Steinerová and Šušol 2007). For instance, children do not consider as much information as adults consider in a buying decision (Winsler et al. 2006). Presenting a lot of product information is therefore not wise if the decision maker is a child. Furthermore, some behavioural variables like self-awareness have been examined in empirical studies (Goukens et al. 2006). Some studies of the same behavioural variables have presented contradictory results due to different operationalisations of the variables. Nevertheless, the influence of behavioural variables on the quantity and quality of information gathering is common sense.

### 3.2 Task

The task of gathering information also influences which and how much information considered by a consumer. Task is mainly characterised by two variables, number of alternatives and number of product attributes. For both variables, a plenty of studies has been conducted especially in the 1980s. Results of these studies are converging. The more alternatives or attributes are available for a consumer, the higher is the absolute
information use and the lower is the relative information use (Lussier and Olshavsky 1979; Capon and Burke 1980). When designing recommender systems, the number of alternatives and attributes considered in average is important in order to provide enough but not too much information to a consumer. Thus, in Table 1 some empirical findings of the use of attributes and alternatives are presented.

<table>
<thead>
<tr>
<th>Product</th>
<th>Number of used attributes</th>
<th>Number of used alternatives</th>
<th>Publication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pocket Camera</td>
<td>5.60</td>
<td>12.00</td>
<td>Sheluga et al. 1979</td>
</tr>
<tr>
<td>Car</td>
<td>8.20</td>
<td>8.80</td>
<td>Ratchford and van Raaij 1980</td>
</tr>
<tr>
<td>Toothpaste</td>
<td>6.58</td>
<td>8.00</td>
<td>Jacoby et al. 1981</td>
</tr>
<tr>
<td>Coffee</td>
<td>4.39</td>
<td>6.20</td>
<td>Knappe 1981</td>
</tr>
<tr>
<td>Camera</td>
<td>6.12</td>
<td>6.31</td>
<td>Knappe 1981</td>
</tr>
</tbody>
</table>

Table 1: Use of product information

Another insight of these studies is the perception that the decision quality decreases if the amount of information presented to the consumer increases her capacity of digesting information (Jacoby 1977, Jacoby 1984). In average consumers are not able to reliably evaluate more than 30 items (Jacoby et al. 1974; Green and Srinivasan 1978).

3.3 Situation

In contrast to early mentions, the situation also influences the process of gathering information (Kakkar and Lutz 1981). Situation encompasses a plenty of variables such as consumer’s budget, time pressure, influence of other persons and daytime (Belk 1975). For example, consumers who are under time pressure consider less information as consumers which are not under time pressure (Knappe 1981). Furthermore, consumers which are under time pressure use more often key information (such as brand and test results) as consumers which are not under time pressure (Newman and Staelin 1972). Due to operationalisation problems, it is not surprising that some results are contradictory. The influence of situational variables is, however, broadly accepted. In order to deduce requirements for a utility-based recommender system, the consumer’s budget seems to be the most important variable which might limit the utility of products. Since other situational variables only have less influence on the goal of constructing a theory-based recommender system, we abstain from a detailed description.

3.4 Problem Orientation

Problem orientation is a composition of person and task and can be interpreted as experience and knowledge a person already has with the considered products and the perceived risk with the products. It seems plausible that the amount of gathered information will decrease if the knowledge and experience with products is growing (Chestnut and Jacoby 1977). This hypothesis is particular supported in some studies, though (Kuss 1987). Experienced consumers evaluate alternatives mostly product by product whereas inexperienced evaluate alternatives mostly attribute by attribute (Bettman and Park 1980). According to Kroebel-Riel and Weinberg (2003), there is a positive relationship between the product involvement and the intensity of gathering
information in the buying decision process. Furthermore, the hypothesis that the higher the perceived risk, the more information are collected is also well proven (Cox 1967; Payne et al. 1993).

3.5 Information Offer
This determinant encompasses variables such as the information source, the format and the design of information as well as costs to gather the required information. The information source and especially the trust in this source have a significant influence on the intensity of gathering information. Consumers typically trust more in tests from independent institutes as in test from any producer. The information format defines whether product information are presented product by product, attribute by attribute or mixed. Bettman and Kakkar (1977) have shown that the information format influences the strategy of gathering information (product by product or attribute by attribute). Most consumers prefer gathering information product by product (Kuss 1987). Furthermore, the hypothesis that the higher the costs for gathering information, the less information are gathered has been successfully confirmed (Silberer and Frey 1981; Punj and Staelin 1983).

4 Design Requirements of Utility-based Recommender Systems
Empirical findings from several studies about the information behaviour have been presented in the previous section. These findings are interesting for creating processes in which consumers must gather, store and digest information. Utility-based recommender systems require evaluations of either entire products or product attributes. Consumers hence have to gather, store and digest information in order to reliably evaluate those attributes or products. Some of the findings presented above can be transformed into design requirements whereas other findings are not appropriate for such requirements. We deduce design requirements in the next subsection and present a prototypical implementation of these requirements in section 4.2.

4.1 Deducing Design Requirements
Personal variables, like the age or the gender, influence the amount and the strategy of gathering information (see section 3.1). Thus, the amount of information should be adaptable according to each consumer’s behaviour. An automatically adaption seems not wise due to the amount of input (age, gender, self-awareness) required to estimate an ideal amount of information.

The number of information gathered from a consumer depends on the number of available information (determinant task) (see section 3.2). In order to ensure reliable attribute or product evaluations, a consumer must have the possibility to consider at least the amount of information presented in Table 1. Furthermore, the number of attributes or products a consumer has to evaluate for eliciting her preferences must be limited up to 30 (see section 3.2).

Situational variables like time pressure are considerable if the consumer can select which and how much information she wants to view (see section 3.3). If the number of information is adjustable, the consumer can apply the recommender system in different situations. Considering the budget of a consumer as a special situational variable is possible if the price of recommendable products can be restricted by the consumer.
In a similar manner, variables of the determinant problem orientation are considerable (see section 3.4). The amount of information must be adjustable according to the experiences and the knowledge a consumer has about products. It seems furthermore meaningful to allow consumers to restrict attribute levels due to avoid information overload.

In order to reduce the risk of a mispurchase, we suggest to additionally present information about the relationship between those attributes a consumer has selected for evaluation or which are used to describe the products a consumer must evaluate.

The effort to gather information can be reduced when information of several products and different attributes are stored in a structured form in the database of the product recommender system. The description of products must also consist of key attributes which are used by experienced consumers as well as consumers which are under time pressure.

Since consumers mostly gather information alternative by alternative (see section 3.5), we suggest to present information product by product in the step of eliciting consumer preferences. Specifying preferences attribute by attribute is thus not recommendable. We suggest applying conjoint analyses or discrete choice analyses which are based on evaluations of entire products and hence allow gathering information product by product.

Table 2 presents a summary of all deduced requirements. In the next subsection, we present a prototypical recommender system which implements all of these requirements.

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>RQ1</td>
<td>Consumers must be able to determine how much and which information they need to build and explicate their preferences.</td>
</tr>
<tr>
<td>RQ2</td>
<td>Consumers must be able to specify attribute level restrictions.</td>
</tr>
<tr>
<td>RQ3</td>
<td>The recommender system must provide enough information for building consumer preferences.</td>
</tr>
<tr>
<td>RQ4</td>
<td>The recommender system must provide key information.</td>
</tr>
<tr>
<td>RQ5</td>
<td>Consumers should evaluate not more than 30 attributes or products for eliciting their preferences.</td>
</tr>
<tr>
<td>RQ6</td>
<td>Consumers should get only those information they need to specify their preferences.</td>
</tr>
<tr>
<td>RQ7</td>
<td>Consumers must be able to understand about attribute relationships.</td>
</tr>
<tr>
<td>RQ8</td>
<td>Consumers should gather information product by product.</td>
</tr>
</tbody>
</table>

Table 2: Requirements for a utility-based recommender system

4.2 Prototypical Implementation
To evaluate the design requirements deduced in the previous section, we implemented a prototype which is using conjoint analysis as preference elicitation method. Necessary statistical adaptations to use conjoint analysis as the core of a utility-based recommender system have been already presented in (Scholz 2008). A comprehension of the system according to the requirements presented in this paper is given in Table 3.
As shown in Table 3 all requirements deduced above are implemented in the prototype. In order to evaluate the prototype, each of these requirements has to be transformed into a hypothesis (Gehlert et al. 2009). In this paper, we present an evaluation of the whole prototype and hence of the interaction of all of the deduced requirements.

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
<th>Requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Product Description Storage</td>
<td>Product descriptions of more than 100 products per product category (notebook, digital compact camera) are stored with key information and more than 50 other information.</td>
<td>RQ4</td>
</tr>
<tr>
<td>Selection of Attributes and Specification of Restrictions</td>
<td>Consumers can select up to 5 attributes for a conjoint analysis and furthermore specify for as many attributes as they want a restriction (minimum, maximum or ideal level). When selecting a particular attribute for the conjoint analysis the relation to other attributes is presented textually.</td>
<td>RQ1, RQ2, RQ3, RQ6, RQ7</td>
</tr>
<tr>
<td>Stimuli Generation</td>
<td>The system generates a set of maximal 16 stimuli (imaginary products) based on Addelman’s basic plans.</td>
<td>RQ5</td>
</tr>
<tr>
<td>Stimuli Evaluation</td>
<td>Consumers must rank the stimuli generated in the previous step according to their preferences.</td>
<td>RQ8</td>
</tr>
<tr>
<td>Utility Estimation and Product Ranking</td>
<td>The system computes the utility of each product based on the product rankings with an ordinary least squares estimator and presents all products meeting the specified restrictions. The products are ordered according to their utility.</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Steps of the interaction with the prototypical recommender system

5 Empirical Study

We conducted two laboratory experiments – one at the University of Regensburg and the other at the University of Passau – in order to assess the prediction validity of the prototype described in subsection 4.2. Predictive validity is here used as measure for decision making satisfaction and has been evaluated with three statistics – first-hit choice rate, rank correlation, and an adaptation of mean absolute error (MAE). We furthermore measured the selection of attributes and definition of restrictions in order to proof whether RQ3 has been fulfilled by the recommender system. Testing the other requirements as well as their interaction effects calls future research. We separately tested RQ3, due to possibility to analyse the effect of this requirement without constructing different treatments (which would require an increased number of test persons).

We recruited 71 undergraduate students overall. Since both laboratory experiments were equal in their procedure, we don’t distinguish between the two experiments. The sample is purposive and the results are hence not generalisable. The experiments were conducted in computer laboratories. Each test person got access to the recommender system. The experiment itself comprises an introduction into the system and two tasks for each participant. In the first task each test person must find an adequate notebook whereas in the second task the participants searched for digital compact cameras. Half
of the test persons started with task 1 while the others started with task 2. 142 questionnaires have been emitted, but only 134 where filled in accurately and complete.

In the introduction the system as well as the experimental procedure was explained. Afterwards each participant fulfilled the two search tasks and completed a questionnaire after each task to assess the predictive validity. In each questionnaire a participant had to re-rank the first 10 results according to the real ranking of the participant. We hence have two ranking vectors, one of the recommender system ($\hat{y}$) and one of the participant ($y$) which we compared in order to assess predictive validity. The vector $\hat{y}$ has been computed based on the rankings of stimuli using an ordinary least squares estimator (see Table 3). The questionnaire encompassed a question for re-ranking products and some other questions to measure reliability and test whether product knowledge and product involvement have an influence on the predictive validity. A significant influence has not been detected for any of these possible moderator variables.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>First-Hit Choice Rate</td>
<td>0.65</td>
<td>0.48</td>
</tr>
<tr>
<td>Kendall’s $\tau$</td>
<td>0.81</td>
<td>0.96</td>
</tr>
<tr>
<td>Spearman’s $\rho$</td>
<td>0.90</td>
<td>0.96</td>
</tr>
<tr>
<td>Weighted Mean Absolute Error</td>
<td>0.30</td>
<td>0.24</td>
</tr>
</tbody>
</table>

Table4: Predictive validity of the recommender system (n=86)

Compared to studies about recommender systems our obtained correlation values are surprisingly high. Hill et al. (1995) achieved a correlation value of 0.62 for their system. De Bruyn et al. (2008) compared three recommendation methods to traditional conjoint analysis and reported on correlation values between 0.38 and 0.50 for their methods and 0.51 for the conjoint analysis. In the same study the authors reported on first-hit choice rates between 0.48 and 0.59.

Rank correlation was computed using both, Spearman’s $\rho$ and Kendall’s $\tau$. In order to calculate mean or standard deviation of correlation coefficient, we used Fisher’s Z-transformation due to get values following a Gaussian distribution. Correlation coefficients are transformed into Z-values as follows:

$$Z = \frac{1}{2} \cdot \ln \left( \frac{1 + r}{1 - r} \right)$$

The average of the Z-values was assessed by the following equation:

$$Z = \frac{\sum_{j=1}^{J} (k_j - 3) \cdot Z_j}{\sum_{j=1}^{J} (k_j - 3)}$$

Here $j$ is the consumer and $k$ the number of products she has re-ranked. There might be less than 10 products in the result set due to the possibility of each participant to restrict each attribute. As equation 2 shows, only those participants having re-ranked more than
3 products could be considered for assessing the predictive validity with rank correlation coefficients. Thus, the sample size of Table 4 is lower than the number of completed questionnaires.

In some studies MAE has been assessed to estimate ranking accuracy of a recommender system. MAE is, however, only appropriate if each participant has re-ranked the same number of products. Otherwise, MAE of each test person must be weighted before the average over all persons is computable. We hence suggest weighting MAE of each participant according to the maximal error conceivable for the participant. The average weighted mean absolute error (WMAE) over all participants $J$ is computed as follows:

$$WMAE = \frac{1}{J} \sum_{j=1}^{J} \frac{1}{K} \sum_{k=1}^{K} | \hat{y}_k - y_k | \left( \sum_{i=0}^{k-1} | k_j - 2i - 2 | \right)$$

(3)

In section 3.2, we stated that consumers use in average between 4 and 9 attributes to make a buying decision and respectively to build preferences. The conjoint analysis implemented in the prototype only supports up to 5 attributes for the step of measuring preferences, though. Since each test person was able to define as many attribute restrictions as she wants, we hope to increase the number of considerable attributes. For example, if a person wants to have a notebook with Bluetooth, a consideration of this attribute is not necessary if a restriction has already been specified. The number of attributes selected for the conjoint analysis is shown in Figure 4.

![Figure 4: Number of selected attributes (n=134)](image)

Only 10.45% of the test persons selected 5 attributes for the conjoint analysis. The number of considerable attributes was hence enough for most of the participants. In Figure 5, the number of specified attribute restrictions is presented. When cumulating both frequencies, the test persons have considered 5.78 attributes in average which is in line with the findings presented in Table 1.

Requirement 3 is seen as fulfilled, since consumers could consider enough attributes to build preferences and to evaluate products in order to measure their preferences.
6 Conclusion

In this paper, we presented a set of design requirements for utility-based recommender systems. The requirements have been furthermore implemented in a prototype in order to empirically examine the requirements. The prototype has shown promise in laboratory experiments and is therefore suitable to validly predict recommendable products. However, there are some limitations of our work which are discussed in subsection 6.1. Implications of the presented findings are depicted in subsection 6.2.

6.1 Limitations

Since we used a purposive and small sample, the findings of our study are hence not representative. We examined predictive validity by comparing the order of the result set with the explicated order of the test persons. This made us realize that the participants were prejudiced by the order generated by the system and may have created biased data. The real predictive validity might be thus lower than reported in this paper. Since the products used in the experiments are both technical products, we have to proof the proposed prototype with other product categories in further experiments.

The evaluation has been conducted with two different samples. Thus, there was a possibility of distinctions between the two samples which may have caused differences in the predictive validity. In order to avoid differences, we controlled the age and gender of the participants and tested whether the number of attributes selected by the test persons or the attributes used had an influence on predictive validity. Significantly differences have not been found for any of these variables. Significant differences of the predictive validity between these samples have also not been detected. Aggregating the results of both samples seems thus possible.

Since the deduction of the design requirements can be erroneous, it is required to empirically test each requirement. Therefore we have to formulate a hypothesis for each requirement in which a positive relationship between the fulfilment of the hypothesis and the predictive validity is stated. In this study, we only empirically examined the implementation of all requirements and compared the resulting predictive validity to
other recommender systems. Testing the requirements separately is planned in future studies.

6.2 Implications
The findings presented in this article imply an impact for both, consumers and recommender system providers. The question of when to use such a system is vital for consumers. Based on a cost-benefit analysis, we can determine the costs existing for learning and handling the system. However, the system also provides some benefits which may exceed costs if the system is used in buying processes which are critical and risky:

- A multiple number of attributes and attribute levels can be considered simultaneously (multi-criteria decision).
- It is possible to reduce the risk of a mispurchase (risk reduction).
- The system is suitable to provide an overview of existing products (product overview).

A practical application of the prototype presented in this paper is only possible if the information about the products existing in a structured form. For our empirically examination, we manually collected and structured product information. The development of a standard format for product descriptions is hence imperative for our as well as for other utility-based recommender systems. Existing formats like EClass are not appropriate for products purchased in B2C e-commerce.

6.3 Future Research
As already discussed in section 6.1 further experiments are necessary in order to test whether each requirement as a separate effect on the predictive validity. Furthermore, possible interaction effects between the requirements will be on focus of future research. Our findings are based on laboratory experiments. In order to proof external validity of our results, conducting field experiments seems promising in future research. Potential partners are price comparison services such as idealo.com as well as retailer of electronic goods like notebookshop.com.

Since we detected that conducting a conjoint analysis is a very time-consuming process, new developments in conjoint analysis like polyhedral methods (Toubia et al. 2007) should be also investigated in future research on utility-based recommender systems.

References


327


