Mining Associations for Interface Design

Timothy Maciag[1], Daryl H. Hepting[1], Dominik Ślęzak[2], Robert J. Hilderman[1]

[1] Department of Computer Science, University of Regina
3737 Wascana Parkway, Regina, SK, S4S 0A2 Canada
[2] Infobright Inc.
218 Adelaide St. W, Toronto, ON, M5H 1W8 Canada
maciagt@cs.uregina.ca, dhh@cs.uregina.ca, slezak@infobright.com, robert.hilderman@uregina.ca

Abstract. Consumer research has indicated that consumers use compensatory and non-compensatory decision strategies when formulating their purchasing decisions. Compensatory decision-making strategies are used when the consumer fully rationalizes their decision outcome whereas non-compensatory decision-making strategies are used when the consumer considers only that information which has most meaning to them at the time of decision. When designing online shopping support tools, incorporating these decision-making strategies with the goal of personalizing the design of the user interface may enhance the overall quality and satisfaction of the consumer's shopping experiences. This paper presents work towards this goal. The authors describe research that refines a previously developed procedure, using techniques in cluster analysis and rough sets, to obtain consumer information needed in support of designing customizable and personalized user interface enhancements. The authors further refine their procedure by examining and evaluating techniques in traditional association mining, specifically conducting experimentation using the Eclat algorithm for use with the authors' previous work. A summary discussing previous work in relation to the new evaluation is provided. Results are analyzed and opportunities for future work are described.

Key words: Association mining, clustering, rough sets, usability, personalization

1 Introduction

The world wide web is increasingly changing the way consumers browse for and purchase items. Millions of consumers engage in purchasing and consuming goods and services from online stores each day. Given this rapid increase in e-market activities there has been an increased demand for more usable tools that more effectively support the online consumer in formulating satisfying decision outcomes. Design of these systems could incorporate functionality that enables consumers to quickly and easily browse for and retrieve items in which they are interested. Providing enhanced options to customize and personalize the support interface could greatly enrich the consumers online shopping experiences [1]. Modelling consumer decision-making strategies in the design of the user interface may aid in achieving this end.

1.1 Consumer Decision-Making

Consumer research has indicated that consumers generally employ two types of decision-making strategies in their purchasing decisions: compensatory and non-compensatory [2].

Compensatory decision strategies are used when the decision maker applies a strict, fully rationalized thought process based on pre-defined preferences, ratings, or rankings to formulate a final decision [3]. The decision-maker will systematically weigh all possible alternatives in order to form the best possible decision outcome. Compensatory decision strategies have the potential to be quite complex in that: the consumer may not always be an expert in the decision domain, the consumer may not value certain attributes yet need to consider them when they formulate their decisions, the decision outcome may consist of an overabundance of information forcing the consumer to filter through results for wanted information, and/or the consumer may have criteria present in every decision yet they still must specify these value(s) in each decision formulation.

Non-compensatory decision strategies are used when the decision-maker applies *bounded rationality* [4]. Bounded rationality refers to the limitations in the human capacity for reaching fully rationalized decision outcomes (i.e. those decision outcomes that consider all facets of available information as in compensatory strategies). Decision makers will often arrive at a final decision based on *ad hoc* decision strategies using a variety of factors, which include: pre-defined and developing preferences, ratings and rankings (total or subset), the interface design, in addition to others [2, 5].

1.2 Usability in Online Shopping Environments

There has been considerable research into understanding what constitutes a satisfying user interface for online shopping environments. Jedetski et al. [6] discuss that the design of the user interface is paramount in whether or not users have a satisfying experience in such support tools. In terms of developing satisfying user interfaces, providing consumers with enhanced options such as the ability to customize and personalize their user interface will ensure that they have a satisfying shopping experience [1, 7, 8]. Holland et al. [9], describe a technique utilizing methods in association mining to gather user preferences in support of developing personalized user interfaces from online user logs. As well, Li and Kit [10] describe a method to enhance the usability of online support tools by utilizing data mining techniques to mine associated information to design and develop a better link navigational structure of a website. Depending on the amount of data and information that consumers must provide, this task has potential to be a highly complex and time consuming. Maciag et al. [8, 3] describe a technique to reduce the complexity of this task by reducing the amount of consumer information required. The primary idea of their research was to formalize the foundations of a personalization procedure aimed at clustering consumers into groups bearing similar attribute values and product preferences.

2 A Review of Previous Work

In Maciag et al. [8, 3], web-based shopping support tools were developed to conduct a usability evaluation. The authors chose to base their evaluation using a software support tool designed by the United States Environment Protection Agency (US-EPA) that enabled product comparisons between 29 environmentally preferable cleaning products using eight product attributes. Table 1 provides a listing of these attributes and their corresponding values.

Table 1. US-EPA attributes (with abbreviations) and corresponding values

Attribute (abbreviation)	**Values**
Skin Irritation (skin)	exempt, negligible-slight, slight, medium, strong, not reported
Food chain exposure (fce)	exempt, $\leq$ 5000, $\leq$ 10000, $\leq$ 15000, $>$ 15000, not reported
Air pollution potential (air)	N/A, 0%, $\leq$ 1%, $\leq$5%, $\leq$ 15%, $\leq$ 30%, $>$ 30%, not reported
Product contains fragrance (frag)	yes, no
Product contains dye (dye)	yes, no
Product is a concentrate (con)	yes, no
Product packaging made of recyclable paper (rec)	N/A, yes
Product minimizes exposure to concentrate (exp)	N/A, yes, no/small sizes, no

56 participants were recruited to complete a series of tasks on the support tools, which included completion of a questionnaire that asked participants to rank the eight attributes described in Table 1 using a four point scale: *unimportant, somewhat important, important, very important*. In addition participants were asked to select which of the 29 cleaning products they would consider purchasing for personal use. This information was used to develop a procedure to gather consumer information in support of clustering the participants into groups having similar attribute and product preferences.

Figure 1 illustrates the authors' procedure in Maciag et al. [3]. First, the 29 cleaning products were clustered, generating four product clusters. A decision system was constructed, comprised of 16 attributes (based on participant rankings) and one decision attribute (based on the participant product selections and product cluster values). Using the Rough Set Exploration System (RSES) [11], rough set reduction techniques were performed on the decision system. Utilizing the genetic algorithm functionality in RSES, the authors formulated the top ten *reducts* (the set(s) of attributes needed to discern objects (e.g. the 56 participants) in a decision system [12]) for the training set and tested the results using participant data in the testing set. Three of the top ten reducts generated consisted of only two of the 16 attributes and had classification accuracy of 100% and total coverage of 88%. It is important to note that the reason why the total coverage was slightly reduced is that some participants in the testing set could not be classified accordingly based on the reduct attributes generated by the genetic algorithm procedure provided by RSES. Thus, the authors proposed a system design where, upon initialization (i.e. a consumer's first use of the support tool), consumers could be given a reduced questionnaire that would elicit their preferences with respect to the attributes represented in the reducts. Consumers would be placed in the appropriate cluster based on their response and a customizable and personalized user interface would be displayed specific to the cluster group's preferences.

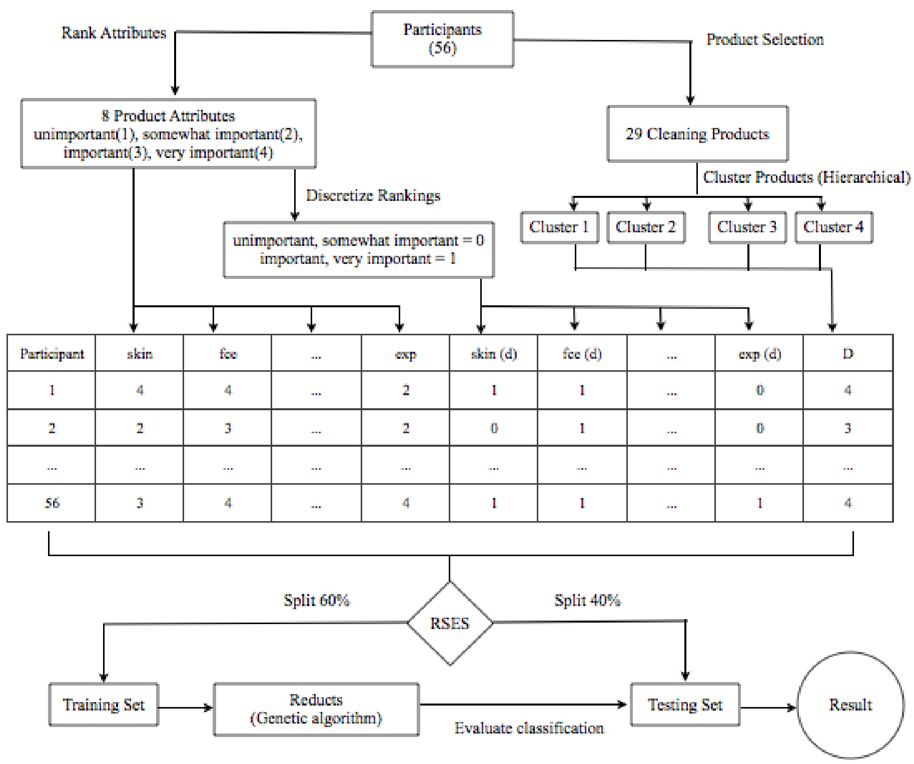

Fig. 1. Diagram illustrating the procedure in Maciag et al. [3]

The work described in this paper will further refine the procedure described in Maciag et al. [3] (Figure 1). Specifically, the research described here will build on previous work by examining and evaluating techniques in association mining to aid in the task of designing the personalized aspects of the user interface after the initial clustering procedure, as described in Maciag et al. [3], is performed. The concepts of compensatory and non-compensatory decision strategies are used to provide the basis for design.

3 Experiment Design and Results

The authors examined and evaluated the Eclat algorithm [13–15], to be used in conjunction with the work described in Maciag et al. [3], as a means to gather useful consumer data and information in support of personalizing aspects of a user interface. Eclat is a data mining algorithm that is used for mining frequent item sets, i.e. sets of *transactions* containing associated values meeting minimum *support* and *confidence* thresholds. These thresholds are described in Equations 1 and 2.

$$Support(X \to Y) = P(X \cup Y) \tag{1}$$

$$Confidence(X \to Y) = \frac{P(X \cup Y)}{P(X)} \tag{2}$$

The Eclat algorithm can be used to determine whether certain items, e.g. item X and item Y, are associated in some fashion [16, 14]. For instance, *what is the percentage items X and Y are purchased together?* [14]. Eclat functions by performing a depth-first traversal of a prefix tree to formulate *association rules*, i.e. the set(s) of rules that could be used to describe relationships among data. Figure 2 provides a classic illustration of the Eclat algorithm [16, 14, 15].

The authors utilized the Eclat algorithm to examine and evaluate the associations among certain aspects of user and product data obtained from the usability evaluation described previously. Figure 3 illustrates the steps taken in the authors' analysis. Eclat software[3] was used to analyze the associations between the total set of 29 cleaning products, the sets of products belonging to each of the four clusters generated in Maciag et al. [3], as well as the stated attribute rankings of those participants assigned within each cluster as per the procedure described in Maciag et al. [3]. A minimum support/confidence threshold of 75% was used in the examination. This support/confidence threshold was chosen since lower thresholds would provide a more loosely bound collection of associated attributes (personalized aspects would potentially loose meaning) whereas as higher thresholds would yield a more tightly bound collection of associated attributes (reduction of design possibilities).

[3] The authors used the Eclat software developed by Borgelt, `http://fuzzy.cs.uni-magdeburg.de/~borgelt/eclat.html`, (Fall 2006) [16]

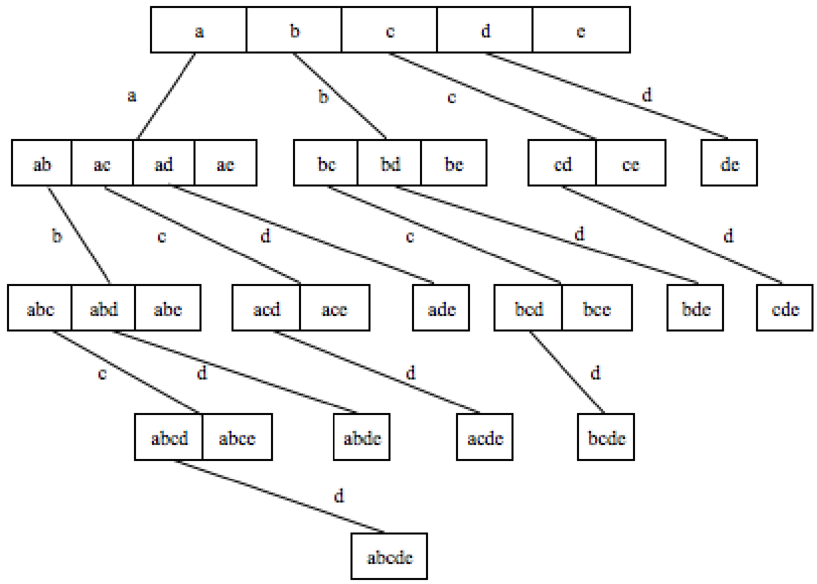

Fig. 2. Classic illustration of Eclat [16, 14]. The empty root is omitted from the illustration. The depth-first traversal begins at the left-most item, *a*, and traverses the tree structure (backtracking when necessary) until all items are analyzed.

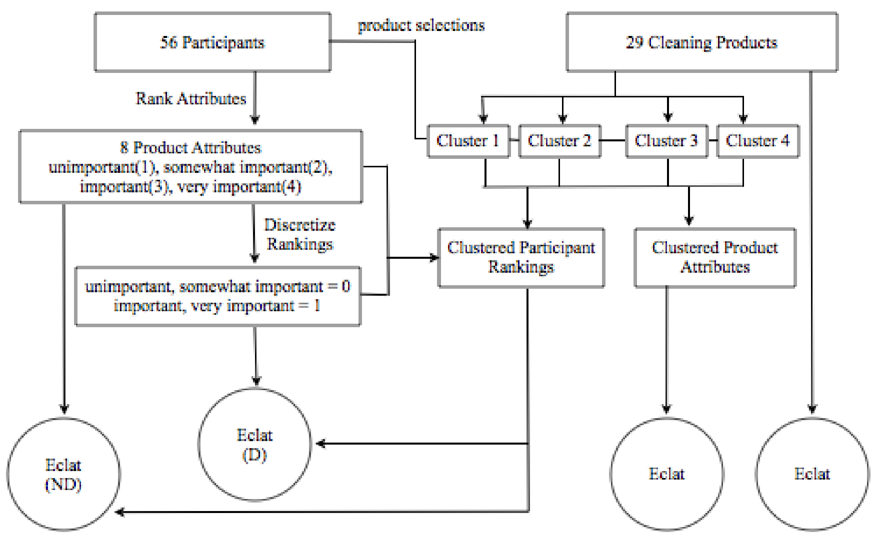

Fig. 3. Diagram illustrating the Eclat procedure used by the authors. Note, on the bottom left hand side, non-discretized attribute rankings (four point ranking scale) are denoted as *ND* and discretized rankings (binary scale) are denoted as *D*

4 Discussion

Tables 2 and 3 provide the results of the authors' examination. The first column of Table 2 represents the cluster value (the total set of products and clusters 1 to 4) as generated by the procedure in Maciag et al. [3], the second column represents the set(s) of associated product attributes generated by Eclat that met the 75% support threshold, and the final column represents the confidence threshold (as %) of the set(s) of associated product attributes. The first column of Table 3 represents the cluster value similarly seen in Table 2, the second column indicates whether the attribute rankings were discretized or not (as illustrated in Figures 1 and 3), the third column represents the set(s) of associated attribute rankings generated by Eclat that met the 75% support threshold, and the final column represents the confidence threshold (as %) of the set(s) of associated attribute rankings.

Table 2. Eclat results for the product analysis (total set and four product clusters). Please refer to Section 2 for definitions and attribute abbreviations. Note there were interesting results below the minimum support/confidence thresholds. These associations are italicized.

Cluster	Associated Attributes	Confidence
All Products	frag=no	86%
1	rec=N/A, dye=no, frag=no, fce=exempt	78%
	dye=no, frag=no, fce=exempt	89%
	fce=exempt	100%
	frag=no	89%
	dye=no	89%
	rec=N/A	78%
2	frag=no, fce=not reported	83%
3	rec=yes, exp=no, con=yes	83%
	exp=no, con=yes	100%
	rec=yes	83%
4	dye=no, frag=no	78%
	fce=exempt, frag=no	78%
	frag=no	89%
	fce=exempt	78%
	dye=no	78%
	dye=no, fce=exempt, frag=no	*67%*

The results described in Table 2 could be used to indicate how to incorporate compensatory decision strategies in the design of a personalized user interface. Using this information, the interface design could highlight those attributes that are highly associated. For example, products in cluster 1 are strongly associated by the following: they contain no dye or fragrance, are not made of recyclable paper, and are mostly exempt from food chain exposure. Since consumers who are

Table 3. Eclat results for the participant rankings. Note, non-discretized (four point ranking scale) are denoted as *ND* under the Type label and discretized rankings (binary scale) are denoted as *D*. Refer to Section 2 for definitions and attribute abbreviations. Note, interesting results below the minimum support/confidence thresholds are italicized.

Cluster	**Type**	**Associated Rankings**	**Confidence**
All Products	D	air=important, skin=important	88%
	D	skin=important	95%
	D	air=important	93%
1	–	*no participants assigned*	–
2	–	*only 1 participant assigned*	–
3	D	exp=not important, rec=not important, con=not important, dye=not important, frag=important, air=important, skin=important	100%
	ND	rec=somewhat important, con=somewhat important, air=very important	100%
4	*D*	*rec=important, air=important, skin=important*	*74%*
	D	air=important, skin=important	93%
	D	skin=important	97%
	D	air=important	95%
	D	rec=important	77%
	ND	*skin=very important*	*72%*

assigned to this cluster would normally select products that have these attribute values, they could initially be included in their product comparisons.

The results described in Table 3 could be used to indicate how to incorporate non-compensatory decision strategies in the design of a personalized user interface. Here, the attributes that are highly associated and favourably ranked could be highlighted on the interface display, while omitting all other non-associated and non-favourably ranked attributes. This information could be used to design and deploy a personalized user interface specific to each cluster's attribute and product preferences.

5 Conclusion and Future Work

This paper described refinement of a procedure to design personalized user interfaces for online shopping support tools. The emphasis of the research described was to formalize a procedure to provide information in support of enhancing the functionality of these types of support tools by accommodating the diversity in consumer decision-making. The authors refined previous research by examining algorithms in association mining, specifically examining the Eclat algorithm. The authors illustrated how Eclat could be used, in conjunction with the authors' previous work, to obtain consumer information in support of designing personalized support tools to enhance the user interface design and potentially increase consumer satisfaction while using such tools.

Future work will include software implementation of the procedure described and further examination through usability evaluation. The authors also plan to examine and evaluate the procedure described in this paper in similar application domains that provide consumers with options to compare items using additional product attributes. As well, techniques to develop metrics for evaluating consumer decision accuracy using concepts described in this paper are currently being designed and evaluated.

References

1. Ha, S.H.: Helping Online Customers Decide Through Web Personalization. IEEE Intelligent Systems (2002) 34–43
2. Bettman, J.R., Luce, M.F., Payne, J.W.: Constructive Consumer Choice Processes. Journal of Consumer Research **25** (1998) 187–217
3. Maciag, T.J., Hepting, D.H., Slezak, D.: Consumer Modelling in Support of Interface Design. In Proc. IEEE International Conference on Hybrid Information Technology **2** (2006) 153–160
4. Simon, H.A.: A Behavioral Model of Rational Choice. Economics **69** (1955) 99–118
5. Hsee, C.K., Leclerc, F.: Will Products Look More Attractive When Presented Separately or Together? Journal of Consumer Research **25** (1998) 175–186
6. Jedetski, J., Adelman, L., Yeo, C.: How Web Site Decision Technology Affects Consumers. IEEE Internet Computing (2002) 72–79
7. Eirinaki, M., Vazirgiannis, M.: Web Mining for Web Personalization. ACM Transaction on Internet Technology **3** (2003) 1–27
8. Maciag, T., Hepting, D.H., Slezak, D.: Personalizing User Interfaces for Environmental Decision Support Systems. In Proc. Rough Sets and Soft Computing in Intelligent Agent and Web Technology (2005)
9. Holland, S., Ester, M., Kiessling, W.: Preference Mining: A Novel Approach on Mining User Preferences for Personalized Applications. In Proc. Principles and Practice of Knowledge Discovery in Databases (PKDD) (2003)
10. Li, C.H., Kit, C.C.: Web Structure Mining for Usability Analysis. In Proc. IEEE/WIC/ACM Web Intelligence (2005)
11. Bazan, J., Szczuka, M.: The Rough Set Exploration System. Transactions on Rough Sets 3 (2005) 37–56
12. Pawlak, Z.: Rough Sets, Theoretical Aspects of Reasoning About Data. Kluwer Academic Publishers (1991)
13. Zaki, M.J.: Scalable Algorithms for Association Mining. In Proc. IEEE Transactions on Knowledge and Data Engineering **12** (2000)
14. Mahanti, A., Alhajj, R.: Visual Interface for Online Watching of Frequent Itemset Generation in Apriori and Eclat. In Proc. IEEE International Conference on Machine Learning and Applications (2005)
15. Ceglar, A., Roddick, J.F.: Association Mining. ACM Computing Surveys **38** (2006)
16. Borgelt, C.: Implementations of Apriori and Eclat. In Proc. Workshop on Frequent Item Set Mining Implementations (FIMI) (2003)