Evaluating the Utility of Web-Based Consumer Support Tools Using Rough Sets

Timothy Maciag[1], Daryl H. Hepting[1], Dominik Ślęzak[2,1], 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. Many popular e-commerce sites provide decision support tools to assist potential customers. Preliminary research indicates that web usage mining analyses may help to assess the utility of these tools and highlight possible areas for improvement. This paper describes a new procedure for assessing the utility of web-based support tools using techniques in rough set theory. The authors evaluated this procedure in a study of two such support tools, one developed by the US-EPA and the other developed by one of the authors. Results provided interesting insights on the utility of both tools and indicated that both tools could be improved. Details of the new procedure, results obtained from its evaluation, and its implications for future work are described.

1 Introduction

Each day millions of consumers connect to the World Wide Web to conduct a variety of e-commerce activities. As such, many current and popular e-commerce websites attempt to provide their consumers with enhanced support tools. As many of these tools evolve, competition to provide the best possible support, thus the best possible shopping experience, intensifies [1]. However, the question remains [2]: How do we evaluate the *utility*, or *usefulness* of these tools? Techniques in web mining, specifically web usage mining which aids in discovering interesting user patterns on the web [3], could assist in this task. Evaluating the utility of support tools may provide online retailers with an indication on the success at which consumers are able to conduct their e-commerce activities while using the tools provided to them. From this perspective, utility may be measured based on the success at which consumers are able to find items, while using the supplied tools, that match their specified preferences. Rough set theory [4], which provides techniques for representing uncertainty in *knowledge systems* [5], may be used for this type of analysis.

A fundamental aspect of rough set theory is formulation of positive, negative, and boundary regions. Given a knowledge system $K = (U, R)$, where U is the universe of all considered objects and R is a binary (equivalence) relation over U, we can approximate any $X \subseteq U$ by groups of objects mutually indiscernible with respect to R. The positive region contains groups fully included in X; the negative region includes groups fully disjoint with X; and the boundary region includes the remaining groups of U. Subsets $X \subseteq U$ correspond to the approximation of decision classes, where *decision rules*, derived from positive, negative, and boundary regions, are formulated and used to classify new cases. One of the primary methodologies of rough set theory that incorporates these concepts is attribute reduction, which refers to the process of discovering minimum (reduced) set(s) of attributes that induce R while maintaining minimal boundary regions. Rough set attribute reduction and classification have been researched in a variety of data mining applications over the last several years [6–9]. In this paper, they are used to obtain measurements of classification accuracy and coverage to indicate a measurement of utility for online shopping support tools.

2 Background

Previously, the authors conducted an evaluation of support tools for environmentally preferable purchasing and analyzed results obtained [10]. Two support tools were used, one based on a support tool provided by the United States Environmental Protection Agency (US-EPA) [11], the other based on a tool developed by one of the authors called *cogito* [12]. Both tools enabled comparisons of 29 environmentally preferable cleaning products using eight product attributes, described in Table 1 [13]. 56 participants were recruited for the evaluation from the University of Regina Computer Science Participant Pool [14].

In the preliminary evaluation described in [10] the participants were asked to answer a series of questions on the two support tools described. Participant response times and task scores were measured. Response times were measured to indicate the duration required by each participant in answering the prescribed questions using the different support tools. Task scores were measured to indicate the success at which participants were able to answer the prescribed questions using the different support tools.

Based on results obtained, participants were more time and task effective while using the *cogito* tool [12]. Although these results were encouraging, they explain little with respect to how well consumers would actually perform on these tools for their own purposes. The authors hypothesize that an analysis of utility may have a greater indication of consumer satisfaction and may provide a clearer illustration of the strengths and weaknesses of provided tools.

3 Evaluation Design

As part of the preliminary usability evaluation described in [10] each of the 56 participants were asked to rank the eight product attributes (in Table 1) according to perceived importance using a four point scale: *unimportant, somewhat*

Table 1. US-EPA attributes (with abbreviations) and corresponding values. According to the US-EPA [13], products with exempt or lower attribute values (skin, fce, air), products with no additives (dye and frag), and products that use reduced (con)/recyclable (rec) packaging are preferred.

Attribute (abbreviation)	**Values**
Skin Irritation (skin)	exempt, negligible-slight, slight, medium, strong, not reported
Food chain exposure (fce)	exempt, ≤ 5000, ≤ 10000, ≤ 15000, > 15000, not reported
Air pollution potential (air)	N/A, 0%, $\leq 1\%$, $\leq 5\%$, $\leq 15\%$, $\leq 30\%$, $> 30\%$, not reported
Contains fragrance (frag)	yes, no
Contains dye (dye)	yes, no
Concentrated packaging (con)	yes, no
Recyclable packaging (rec)	N/A, yes
Minimizes exposure to concentrate (exp)	N/A, yes, no/small sizes, no

important, important, very important. Each participant was also asked to select a product they would consider using for personal applications while using either the US-EPA or *cogito* tool. This information was used as part of the criteria to determine a measurement of utility for each tool.

First, the 29 cleaning products were clustered into product groups using hierarchical clustering with the *maximum* distance metric, as described in [15]. From this analysis, four product clusters were found. In order to conduct a proper analysis of product attribute values and specified user rankings, each individual product attribute value needed to be re-coded and mapped to corresponding user rankings. In this preprocessing phase, all data was re-coded into binary. This process is described in Table 2.

A train and test procedure was conducted. The training set was comprised of the 29 cleaning products with re-coded attributes as knowledge system attributes and product cluster membership values as the decision classes. Two testing sets were created and comprised of the participants' re-coded attribute rankings as knowledge system attributes and product selections based on cluster membership values as the decision class (e.g. if a participant selected a product belonging to cluster 4, this value was put as his/her decision class). The testing sets were split into those participants who selected products using either the US-EPA or *cogito* tool. Some participants in the evaluation did not select a product (ten and four participants did not select a product when using the US-EPA and *cogito* tool respectively). These participants were omitted from further evaluation.

Rough set reduction techniques were performed on the training set to reduce the number of attributes required for analysis. This allowed emphasis and focus on only those product attributes absolutely required to discern products based on their cluster associations. Although there are only eight attributes in the

Table 2. Description of the re-coding procedure used in the analysis. See Table 1 for attribute abbreviations and values. Here, product attributes and user rankings are recoded into binary and mapped according to the described ranges to indicate their associative strength. Rankings of important and very important, and product attributes within the ranges described in this table, were re-coded as 1 and 0 otherwise. For example, if a participant ranked skin irritation highly (important or very important, value of "1"), products with lower skin irritation (slight or less, value of "1") would be considered preferred.

	Participant Rankings and Product Value Ranges	
Attribute	**Important/Very Important**	**Not Important/Some Important**
skin	slight of less	medium and stronger
fce	less than 5000	10000 and greater
air	1% or less	5% and greater
frag	no	yes
dye	no	yes
con	yes	no
rec	yes	N/A
exp	no	yes

tools used in the evaluation described, other e-commerce support tools may include a larger number of attributes. Reduction of these attributes may greatly reduce the complexity of the analysis as well as the computational complexity of the described procedure. Results obtained from the reduction procedure were applied to each testing set and obtained classification accuracy and coverage measurements were analyzed.

4 Results and Discussion

Based on the rough set reduction, two reducts, each with five attributes were generated: {skin, fce, air, dye, con} and {skin, fce, air, dye, exp}. Each reduct had a positive region of 86.2%. Since the positive region was not exactly 100%, this may indicate that there exist some cleaning products among the four clusters generated that were difficult to cluster. This may have been the result of the attribute re-coding process described in Table 2 or it could be that although certain products were assigned to different clusters, it may be that the distance between a select few of these products is minimal, e.g. products in opposing clusters could be relatively close in proximity. This analysis is left for future evaluation.

Table 3 describes the results obtained. Upon first inspection, the results indicate the majority of participants selected products residing in one cluster (cluster 4). It may be that a larger, more diverse participant sample (including participants outside the University setting) may yield more inclusive results. However, when analyzing the products residing in product cluster 4, it is interesting to note that these cleaning products are the most environmentally friendly among

Table 3. Evaluation results. Accuracy and coverage measures are used to indicate a measurement of utility.

US-EPA Train/Test Results							
Actual	**Predicted**				**#Objects**	**Accuracy**	**Coverage**
	3	4	1	2			
3	0	0	0	0	1	0	0
4	0	13	0	2	17	86.7%	88.2%
1	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0
Totals					**18**	**86.7%**	**83.3%**
***cogito* Train/Test Results**							
	3	4	1	2			
3	0	0	0	0	1	0	0
4	1	16	0	0	22	94.1%	77.3%
1	0	0	0	0	0	0	0
2	0	0	0	0	1	0	0
Totals					**24**	**94.1%**	**70.8%**

the 29 products provided by the tools. The authors hypothesize that consumers who use this type of tool may be more interested in environmentally friendly selections. Thus, it may be that a higher percentage of consumers would belong to this cluster, as seen by the results in the authors' evaluation.

The authors hypothesize that utility may be analyzed from the perspective of obtained accuracy and coverage measurements. When observing the results for accuracy totals, slightly higher accuracy was obtained by the *cogito* testing set. However, when observing the coverage measurements obtained by the *cogito* testing set, results indicate that there may be a dramatic difference between the preferences selected by participants in the testing set and the product attributes in the training set.

When observing results obtained by the US-EPA testing set, classification accuracy is slightly less than that obtained by the *cogito* testing set. However, coverage is higher indicating that the difference between the preferences selected by the participants in the US-EPA testing set and the product attributes in the training set was less severe than that observed in the *cogito* testing set. In either case, it would seem that each support tool would benefit from improvements made to their interfaces since neither tool obtained complete accuracy or coverage measurements.

5 Conclusion

This paper described a procedure using techniques in rough sets to evaluate the utility of web-based consumer support tools. Based on the knowledge obtained from the experiments described, the results did yield interesting indicators on the overall utility of the tools and indicated that both support tools would benefit from new design considerations or additional design enhancements. Future work will include applying and evaluating the described procedure to similar web-based support tools with additional attributes as well as web-based support tools in other shopping domains.

Acknowledgements The authors would like to acknowledge the support of the Centre for Sustainable Communities at the University of Regina, the Faculty of Graduate Studies and Research at the University of Regina, and the Natural Sciences and Engineering Research Council (NSERC) of Canada.

References

1. Ha, S.H.: Helping Online Customers Decide Through Web Personalization. IEEE Intelligent Systems (2002) 34–43
2. Spiliopoulou, M.: Web Usage Mining for Web Site Evaluation. Communications of the ACM **43** (2000) 127–134
3. Kosala, R., Blockeel, H.: Web Mining Research: A Survey. Special Interest Group on Knowledge Discovery and Databases (SIGKDD) Explorations **2** (2000) 1–15
4. Pawlak, Z.: Rough Sets, Theoretical Aspects of Reasoning About Data. Kluwer Academic Publishers (1991)
5. Tan, R.R.: Rule-Based Life Cycle Impact Assessment Using Modified Rough Set Induction Methodology. Environmental Modelling and Software (EMS) **20** (2005)
6. Goebel, M., Gruenwald, L.: A Survey of Data Mining and Knowledge and Discovery Software Tools. Special Interest Group on Knowledge Discovery and Databases (SIGKDD) Explorations **1** (1999) 20–33
7. Lingras, P., West, C.: Interval Set Clustering of Web Users with Rough K-Means. Journal of Intelligent Information Systems **23** (2004) 5–16
8. Ślęzak, D.: Association Reducts: A Framework for Mining Multi-Attribute Dependencies. In Proc. International Symposium on Methodologies for Intelligent Systems (ISMIS) (2005) 354–363
9. Chimphlee, S., Salim, N., Ngadiman, M., Chimphlee, W., Srinoy, S.: Independent Component Analysis and Rough Fuzzy Based Approach to Web Usage Mining. In Proc. International Conference on Artificial Intelligence and Applications (IASTED) (2006) 422–427
10. Maciag, T.J.: An Evaluation of User Interfaces for Environmental Decision Support Systems. Master's thesis, University of Regina (2005)
11. Hepting, D.H., Maciag, T.: Consumer Decision Support for Product Selections. In Proc. International Symposium on Environmental Software Systems (ISESS) (2005)
12. Hepting, D.H.: Towards a Visual Interface for Information Visualization. In Proc. Information Visualization (IV) (2002) 295–302
13. United States Environmental Protection Agency, U.S.: Cleaning Products Pilot Project Fact Sheet. http://www.epa.gov/oppt/epp/pubs/cleanfct.pdf (1997) (Accessed April 2007)
14. Hepting, D.H.: Ethics and Usability Testing in Computer Science Education. ACM Special Interest Group on Computer Science Education (SIGCSE) **38** (2006) 76–80
15. Maciag, T., Hepting, D.H., Ślęzak, D., Hilderman, R.J.: Mining Associations for Interface Design. In Proc. Joint Rough Set (JRS) Conference (To appear) (2007)