Evaluation of a Dominance-Based Rough Set Approach to Interface Design

Timothy Maciag, Daryl H. Hepting, Robert J. Hilderman
University of Regina
Department of Computer Science
3737 Wascana Parkway, Regina, SK, S4S 0A2 Canada
maciagt@cs.uregina.ca, dhh@cs.uregina.ca, robert.hilderman@uregina.ca

Dominik Ślęzak
Infobright Inc.
218 Adelaide St. W, Toronto, ON, M5H 1W8 Canada
dominik.slezak@Infobright.com

Abstract

This paper explores refinements to methods used in a procedure being developed by the authors to personalize user interfaces for online shopping support tools. In the authors' original procedure, classical methods in rough set theory are used in conjunction with traditional algorithms in web usage mining. This paper will explore an alternative approach, specifically the dominance-based rough set approach (DRSA), for use with the authors' original procedure. DRSA has its foundations in the classical rough set approach (CRSA). However unlike CRSA, DRSA considers feature/preference-ordered data. In web usage mining analyses, where elicitation of user preferences is a common task, feature/preference order is an important factor and may provide insights that classical/traditional approaches may omit. The authors discuss how DRSA may benefit and improve their original procedure and discuss how the information gained from DRSA analyses could be used to further build their original procedure by enabling item ordering and feature highlighting. This paper will describe the research process, outcomes, and outline opportunities for future work.

1 Introduction

Many e-commerce sites provide tools for consumers that facilitate their online shopping activities. Some of these tools, e.g. such as the ones provided by Amazon [1], provide customizable and personalized displays and functionality with the goal of increasing consumer satisfaction. These personalized displays are often designed using knowledge obtained from web usage mining analyses [14]. Consumer preferences may be elicited and previous consumer transactions analyzed with the goal of attaining useful knowledge about consumer habits. Often this requires that consumers take time away from their shopping activities to input substantial information about their preferences and transactions, which goes against Raskin's [16] *laws of user interface design*, one of which being *a computer should not waste the user's time nor require them to do more work than necessary*.

CRSA, introduced by Pawlak [15], and traditional data mining algorithms, such as Apriori or Eclat [3], could assist in conducting these analyses and in developing a procedure to design, develop, and deploy an effective and highly efficient online shopping support system quickly so that consumers can benefit from enhancements sooner, without having to spend time developing their profiles [13]. However, CRSA and Apriori/Eclat may not always provide the *best* results, as consideration of preference-ordered data is not taken into account. In web usage mining analyses, where elicitation of user preferences is a common task, feature/preference order is an important factor and may provide further insights into new ideas for personalized designs. Alternatively, DRSA, proposed by Greco et al. [4, 6, 18, 7] does provide methods that consider feature/preference order and may provide a more adequate approach when conducting analyses from this perspective.

1.1 Eclat, CRSA, and DRSA

Eclat [20, 2] is a highly effective traditional data mining algorithm that could be used for obtaining useful consumer information. Specifically, Eclat is used for mining

---

frequent item sets, sets of transactions, or items (e.g. products), containing associated values meeting minimum support and confidence thresholds. For example, the percentage (%) item *x* and item *y* are purchased together. Eclat is usually used to observe associations among data from all facets, i.e. generally all features are analyzed together to determine their true associative strengths with each other. Figure 1 illustrates the functionality of the Eclat algorithm.
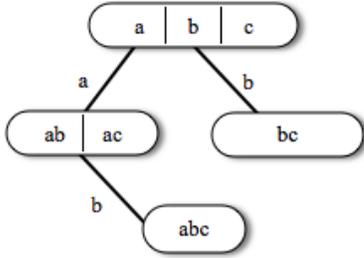

Figure 1. Illustration of Eclat. Here, a depth-first prefix tree traversal of 3 items, a, b, c, **is performed [2]**

CRSA provides methods to reduce the complexity of data through analysis of data equalities, data is either *equal* or *not equal*. In CRSA, reduction methods are performed based on *positive, negative, and boundary regions* of classes represented in a *knowledge system*. For instance, consider a knowledge system $K = (U, R)$, where U comprises all objects in a *knowledge space* and R is an *equivalence relation* over U [12]. $X \subseteq U$ can be approximated by groups of objects mutually indiscernible with respect to R. In this sense, those groups fully included in X comprise the *positive region*, those groups fully disjoint with X comprise the *negative region*, and all remaining groups comprise the *boundary region*. Subsets $X \subseteq U$ correspond to the approximation of decision classes, where *decision rules*, derived from these three regions, are formulated and used to classify new cases. This is performed with the goal of evaluating the possibility of feature reduction, which refers to the process of discovering minimum set(s) of features, called *reducts*, that induce R while maintaining minimal boundary regions.

DRSA is similar to CRSA, however, the approach is based on *dominance relations*, e.g. inequalities among data, as opposed to *indiscernibility relations*, e.g. equalities among data [5]. Similarly to the example of CRSA, consider a knowledge system $K = (U, R)$, where U comprises all objects in the knowledge space. Here, objects $\in K$ are comprised of *gain* features ($\uparrow$), where higher/increasing values are preferred, and/or *cost* features ($\downarrow$), where lower/decreasing values are preferred. R represents a *dominance relation* over U where objects are evalu-

ated as being *equal to, or more preferred* ($\geq$) or *equal to, or less preferred* ($\leq$) other objects in K based on their respective decision class designations [5]. Positive, negative, and boundary regions are formulated based on R. For instance, for objects x and $y \in U$, if all features in object x, including the decision attribute D_x, are *equal to, or more preferred* to all corresponding features in y, x is said to *dominate y*. Here, x would be in the positive region for the *dominance relation* $\geq D_x$ whereas as y would be in the positive region for the *dominance relation* $\leq D_x$.

1.2 CRSA vs. DRSA: Illustration

Consider a knowledge system comprised of environmentally friendly cleaning products, as described in Table 1. Illustrated are five contrived cleaning products, with values for three environmental features, *skin irritation, air pollution percentage, and recyclable packaging*, and one decision attribute, *Environmental Protection Agency (EPA) rating*. Preferred products include those that have lower skin irritations (none < mild < medium < strong), lower air pollution percentages, use recyclable packaging, and have higher contrived EPA ratings.

Table 1. Example of a knowledge system for comparison of CRSA and DRSA.

	Skin irritation	Air pollution	Recyclable packaging	EPA rating
a	none	$\leq 1\%$	yes	4
b	none	$\leq 3\%$	yes	3
c	medium	$\leq 5\%$	no	2
d	strong	$\leq 10\%$	no	1
e	mild	$\leq 1\%$	yes	4

CRSA would generate positive, negative, and boundary regions based on each specific EPA rating. For example, referring to Table 1, the positive regions for an EPA rating of *4* would include products *a* and *e*; an EPA rating of *3* would include product *b*; an EPA rating of *2* would include product *c*; and an EPA rating of *1* would include product *d*. CRSA feature reduction methods could also be applied. It is important to remember that CRSA analyzes data equalities, i.e. *equal* or *not equal* comparisons. When analyzing Table 1, our goal is to reduce the number of features needed to satisfactorily discern products based on their assigned EPA ratings. If we first attempt to remove *skin irritation*, we observe that we can still satisfactorily discern products based on their respective EPA ratings. If we remove *skin irritation* and *air pollution*, we can no longer discern products based on their respective decision classes since products *a*, *b*, and

e have the same value for *recyclable packaging* yet products {*a*, *e*} and {*b*} have conflicting EPA ratings. If we remove *skin irritation* and *recyclable packaging*, we still can satisfactorily discern products based on their EPA ratings. Since we require at least one of the three features to discern products based on their EPA ratings, we can stop here and remove the features *skin irritation* and *recyclable packaging*. Thus, *air pollution* comprises a reduct and as such is the only feature required in the analysis to discern products based on their EPA ratings.

DRSA would generate positive, negative, and boundary regions based on dominance-based relations between the features and the EPA ratings. For example, the positive regions for products assigned EPA ratings *at most* (≤) *2* would include products *c* and *d*; and *at most 3* would include products *b*, *c*, and *d*. Similarly, the positive regions for all products assigned EPA rating *at least* (≥) *2* would include products *a*, *b*, *c*, and *e*; an EPA rating of *at least 3* would include products *a*, *b*, and *e*; and an EPA rating of *at least 4* would include products *a* and *e*. As in CRSA, DRSA feature reduction methods can be applied. However, when observing the information provided in Table 1 and the results described for the positive, negative, and boundary regions, it is observed that no reducts would be generated since all features are required to describe the dominance-relations among the EPA ratings. For example, all products that have no skin irritation and are manufactured using recyclable packaging would be assigned an EPA rating of *at least 3*, these being products *a* and *b*. As well, all products that have mild skin irritation and are manufactured using recyclable packaging would be assigned an EPA rating of *at least 4*, these being products *a* and *e*. In order to determine the *dominance-based* relations of products *a*, *b*, and *e*, we also need to observe the *air pollution* percentage of each product. Thus, no features can be reduced.

The choice between CRSA and DRSA depends on the nature and interpretation of the data in the analysis. Although in CRSA, discretization [1] techniques could be employed to derive preference-order, DRSA may be preferred since this additional pre-processing step could be omitted.

The rest of this paper is organized as follows. Section 2 provides a summary of our original procedure using CRSA and traditional data mining algorithms. Section 3 describes an empirical evaluation of our original procedure using DRSA followed by a discussion of results obtained from the evaluation in Section 4. We conclude in Section 5 with a few remarks and suggestions for future research.

2 Summary of Previous Work

We previously designed a procedure using CRSA and Eclat to obtain useful knowledge from data acquired in a recent usability evaluation in pursuit of personalizing user in-

Table 2. US-EPA features (with abbreviations) and corresponding values. Integer values in brackets represent the scaling used in the analysis. Lower integer values are preferred.

Feature (abbr.)	Values
Skin irritation (skin)	exempt (0), negligible-slight (1), slight (2), medium (3), strong (4), not reported (5)
Food chain exposure (fce)	exempt (0), ≤ 5000 (1), ≤ 10000 (2), ≤ 15000 (3), > 15000 (4), not reported (5)
Air pollution potential (air)	0% (0), ≤ 1% (1), ≤5% (2), ≤ 15% (3), ≤ 30% (4), > 30% (5), not reported (6)
Fragrance (frag)	no (0), yes (1)
Dye (dye)	no (0), yes (1)
Concentrated packaging (con)	yes (0), no (1)
Recyclable paper (rec)	yes (0), N/A (1)
Minimizes concentrate exposure (exp)	no (0), no/small sizes (1), yes (2)

terfaces for online shopping support tools. As basis for our analysis a tool developed by the Unites States Environmental Protection Agency (US-EPA), to conduct comparisons and product selections of environmentally preferable cleaning products, was used. The product database was comprised of 29 cleaning products, each with values for eight environmental features (attributes). These are described in Table 2 [19].

56 participants were recruited from the University of Regina Computer Science Participant Pool [8] and asked to perform tasks on the tools and answer questions concerning their purchasing habits. Of the 56 participants, only 42 participants provided complete questionnaires. The information for these 42 participants formed the empirical foundations for the procedure designed [10, 11, 13].

Our original procedure was split in two phases, a consumer clustering phase (phase 1) and an association mining phase (phase 2). In the first phase, the 29 cleaning products were clustered into four product groups using hierarchical clustering. Participants were asked to rank the eight features using a four point scale (*unimportant (1), somewhat important (2), important (3), and very important (4)*)[2] and

---

[2]Note, the integer values in brackets correspond to the scaled values
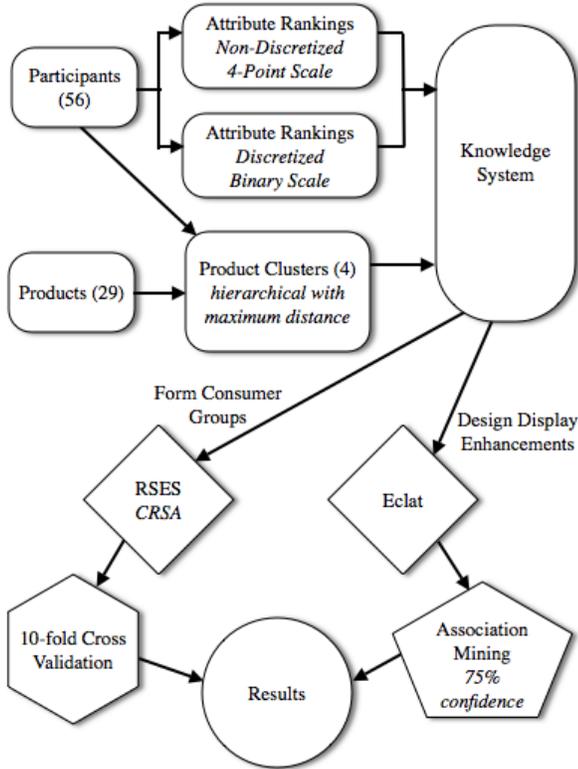
Figure 2. Model-View of our procedure using the previous methods described. Discretization of features was conducted. RSES [1] software and Borgelt's Eclat algorithm [2] were used in the analysis.

Table 3. Formulated Reducts (Top ten) using CRSA. Features additionally labelled with "$_d$" refer to the discretized rankings

Num	Size	Reducts
1	2	*dye, con*
2	2	*fce, rec*
3	2	*fce, con*
4	2	*skin, exp*
5	2	*fce, exp*
6	2	*air, con*
7	3	*frag, dye, exp*
8	2	*rec, fce$_d$*
9	2	*con, fce$_d$*
10	2	*con, dye$_d$*

select one of the 29 cleaning products they would consider purchasing for personal use.

Manual (expert-based) discretization of the participant's feature rankings was conducted to account for the discrepancies between the different ranges as it may be quite difficult to fully understand the difference between the ranking designations. For instance, some consumers may consider features ranked as *important* or *very important* similarly, with only minor favour towards features ranked as *very important*. However, it may be that some consumers place higher value on features ranked as being *very important*. Given the unclear nature of the ranking scheme, a binary scale was coded from the participants' four-point rankings to distinguish features ranked as *important or greater (3-4)* versus all others *(1-2)* to account for these discrepancies. We differ from other approaches by keeping both discretized and original features for further analysis. Since in CRSA, the reducts are calculated to include both dis-

cretized and non-discretized versions of the same feature, we may say that a minimal consumer-related information is searched concurrently at two levels, i.e. the choice of features, as well as the choice between more general and more detailed information provided by those features. A knowledge system was constructed for analysis, comprised of 16 features (non-discretized and discretized values) and one decision class corresponding to each participant's product selection based on its associated product cluster value *(1-4)*.

The goal of the first phase was to use CRSA to discover the minimal amount of consumer information required to discern contrasting consumer groups [10, 11]. The primary idea was to observe the possibility of reducing the amount of information required by consumers before personalized user interfaces are theoretically possible [11, 13]. Results obtained in our original procedure are given in Table 3. Here, the top ten reducts are generated using the genetic algorithms functionality provided in RSES [1] (used to capture and produce the *n best possible* reducts) are displayed. We evaluated classification accuracy and coverage using a 10-fold cross validation method [9] performed ten times averaging the results to reduce variability. Results concluded that a global classification accuracy and coverage of 87.5% and 100% were respectively obtained.

The goal of the second phase of our procedure was to discover associations, using the Eclat algorithm, between the products residing in each cluster and the rankings of the participants assigned to each cluster. It is important to note that all features were used in this phase and not only the features represented in the calculated reducts in phase 1. The decision to include all features and not only the reduct features was made to provide a more detailed view of the product features and feature ranking associations respective in

---

used in the described analyses.

each cluster. It was hypothesized that by only focusing on a partial set of features, such as only those features comprising the reducts, a true depiction of the associations among product features and consumer feature rankings, may not be possible. However, this is left for future evaluation. As well, only the eight non-discretized features were analyzed in the evaluation of Eclat. It may also be interesting to evaluate the discretized features alongside the non-discretized ones as it may provide additional information. This is also left for future evaluation.

The primary idea of the second phase of our procedure was to gather information in support of designing enhanced search functionality to aid consumer decision-making strategies by designing display aids for *compensatory* (fully rationalized) and *non-compensatory* (*bounded rationality* [17]) decision-making processes [13]. Compensatory decision-making strategies are used when consumers fully rationalize their decision, evaluating all features regardless of their perceived importance. Non-compensatory decision-making strategies are used when consumers choose a smaller subset of features, those which may be perceived as being more important, to formulate their decisions. The general idea for the knowledge obtained in our second phase was to highlight those features that are highly associated based on product features and consumer rankings respective within each cluster, enabling consumers to formulate decisions more efficiently and effectively [13]. A support and confidence threshold of 75% was used in the analysis, with some exceptions being relatively close to 75% [13]. Results obtained from this analysis are presented in Tables 4 and 5 [13].

Results illustrated in Table 4 describe associations among products residing in respective clusters. Here, the idea is for product feature associations to be highlighted to incorporate support for compensatory decision-making strategies. For example, since the majority of products in cluster 3 are manufactured using concentrated recyclable packaging, these values could be pre-included in the consumer's search processes since it is highly probable that these consumers may have these preferences. Results illustrated in Table 5 describe associations among the eight feature rankings. The idea here is for these to be highlighted to incorporate support for non-compensatory decision-making strategies. For example, since the majority of consumers in cluster 4 ranked recyclable packaging, air pollution potential, and skin irritation as important, these features could be highlighted in both the consumer's search processes and the resulting product descriptions.

3 Evaluation Using DRSA

As mentioned previously, DRSA may provide a more realistic depiction of consumer preferences and product sim-

Table 4. Eclat results for the *product* analysis. Associations above 75% were observed. Interesting associations under this threshold are italicized. Refer to Table 2 for integer value designations. Lower integer values are preferred (environmentally)

Cluster	Associations	Conf
1	rec=1, dye=0, frag=0, fce=0	78%
1	dye=0, frag=0, fce=0	89%
1	fce=0	100%
1	frag=0	89%
1	dye=0	89%
1	rec=1	78%
2	frag=0, fce=5	83%
3	rec=0, exp=0, con=0	83%
3	exp=0, con=1	100%
3	rec=1	83%
4	dye=0, frag=0	78%
4	fce=0, frag=0	78%
4	frag=0	89%
4	fce=0	78%
4	dye=0	78%
4	*dye=0, fce=0, frag=0*	*67%*

ilarities than results obtained when using CRSA and Eclat (and other traditional data mining algorithms) since DRSA analyzes data in terms of inequalities, which may be considered more *natural* with respect to the consumer decision-making process [17].

In our original procedure, the order of the product clusters was originally not taken into consideration, i.e. we did not determine if one cluster was comprised of *better* products than another. Since DRSA considers preference-order, we deemed it necessary to develop a procedure to order the product clusters as such. The task of ordering product clusters has potential to be complicated as it may be difficult to decide which products are *better* than others since this type of preference analysis is very much consumer-specific, e.g. certain product features may have more bearing on a consumer's decision formulation given their preference sets. Given this inconsistency, a more broad analysis of the products needed consideration. As such, we devised a procedure to order the product clusters, from *worst* to *best*, evaluating each product feature as being equal. For this procedure, the following steps were taken. First, mean values for each product feature were calculated specific to products assigned to each cluster. Secondly, the mean values for each *cost* feature were added together, subtracting the mean value for each *gain* feature. Based on results obtained, prod-

Table 5. Eclat results for the *rankings* **analysis. No results were obtained from clusters 1 and 2. Interesting results below the 75% confidence threshold are italicized. Non-discretized rankings are denoted as** *ND***, discretized as** *D***. In the discretized scaling, rankings of unimportant and somewhat important are denoted as 0 whereas rankings of important and very important are denoted as 1.**

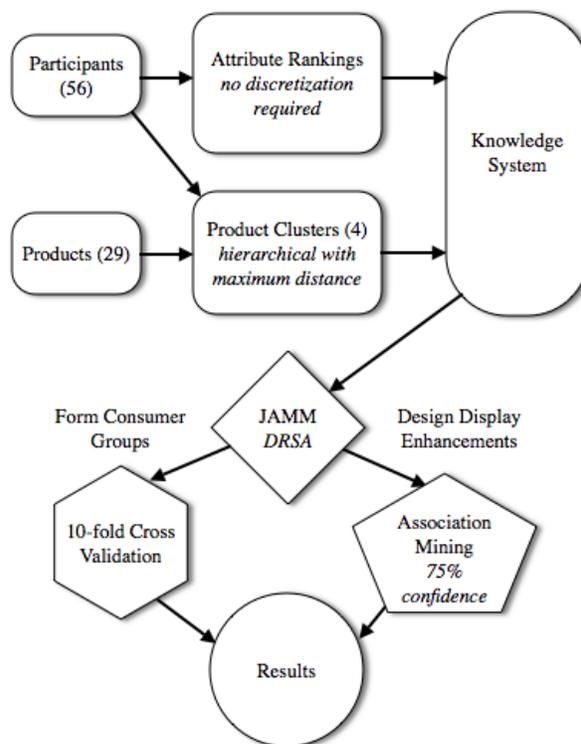| Cluster	Type	Associations	Conf
3	D	exp=0, rec=0, con=0, frag=1, dye=0, air=,1 skin=1	100%
3	ND	rec=2, con=2, air=4	100%
4	*D*	*rec=1, air=1, skin=1*	*74%*
4	D	air=1, skin=1	93%
4	D	skin=1	97%
4	D	air=1	95%
4	D	rec=1	77%
4	*ND*	*skin=4*	*72%*

Figure 3. Model-View of our procedure using DRSA. In the new method, discretization of features was omitted and the new product cluster ordering was applied. As well, the JAMM software was used for the DRSA analysis of both the consumer grouping and association mining.

uct clusters were re-ordered accordingly.

Figure 3 provides a model view of our evaluation using DRSA. It is similar to our original procedure illustrated in Figure 2, however, somewhat more simplified. The JAMM[3] software application was used for the evaluation. The first task in the analysis of DRSA was to re-evaluate the first phase in our original procedure, the consumer clustering phase. In the new evaluation, the discretization of the ranked preference data was omitted since DRSA accounts for the inconsistencies of preference-ordered data. Thus, the knowledge system was only comprised of the eight features as opposed to the 16 in the original evaluation (discretized and non-discretized rankings). Reduct analysis was conducted. Results obtained are described in Table 6.

Similarly to our previous testing method, a 10-fold cross validation method was used (again, performed ten times to reduce the high variability of results) to test the newly generated DRSA reducts and their classification accuracy. Results concluded that an average global classification accuracy of 85.7% with a 100% average global coverage was obtained. These results were only slightly lower than those obtained in our original evaluation using CRSA (87.5% and 100% accuracy and coverage).

It is interesting to note that reducts are smaller when us-

---

[3]JAMM is an application for conducting DRSA analyses. It was developed by the Laboratory of Intelligent Decision Support Systems at Poznań University of Technology under the management of Roman Slowinski. It is freely available online for non-commercial applications at: http://www-idss.cs.put.poznan.pl/site/jamm.html.

ing CRSA. When using DRSA, the larger sized reducts and the slightly lower classification accuracy may likely be due to the *richer* decision characteristics in the case of DRSA in relation to CRSA. In CRSA, data can either equal or not equal whereas in DRSA, dominance-based relations need consideration. Thus, in DRSA, it may be more difficult to approximate decision outcomes.

The second task was to re-evaluate the second phase of our original procedure, i.e. the association mining phase. As in our original evaluation [13], associations meeting a 75% support and confidence threshold were observed, with some exceptions being *relatively close* to the 75% threshold. Results from the DRSA analysis are provided in Tables 7 and 8. The associations illustrated for both the product-view, Table 7, and the ranking-view, Table 8, were quite unique from results obtained in the previous study using Eclat. We hypothesize that newly obtained results may complement previously obtained results by providing ad-

Table 6. Formulated reducts using DRSA. *Core* **features, those features in every reduct, include** {**dye, rec**}

Num	Size	Reducts
1	4	*air, dye, rec, exp*
2	4	*fce, air, dye, rec*
3	4	*frag, dye, rec, exp*
4	4	*fce, frag, dye, rec*
5	3	*dye, con, rec*

Table 7. DRSA results for the *product* **analysis. Refer to Table 2 for integer value designations. Lower integer values are preferred (environmentally). Italicized associations refer to those under the 75% confidence threshold but still considered interesting.**

Cluster	Associations	Conf
≤ 2	air ≥ 3	78%
≤ 3	skin ≥ 3	80%
≥ 3	air ≤ 1	73%
≥ 4	skin ≤ 2, air ≤ 1	78%
≥ 4	fce ≤ 2, air ≤ 1	78%
≥ 4	*skin ≤ 2, fce ≤ 2, rec ≤ 0*	67%
≥ 4	skin ≤ 2, air ≤ 2, dye ≤ 0	78%
≥ 4	*fce ≤ 2, air ≤ 2, rec ≤ 0*	67%

Table 8. DRSA results for the *rankings* **analysis. Unimportant (1), somewhat important (2), important (3), and very important (4). Italicized associations refer to those under the 75% confidence threshold but still considered interesting.**

Cluster	Associations	Conf
≤ 3	*{skin ≤ 3, exp ≤ 1}* or *{fce ≤ 2, dye ≤ 2, exp ≤ 1}* or *{dye ≤ 2, con ≤ 2, exp ≤ 1}* or *{dye ≤ 2, rec ≤ 2, exp ≤ 1}*	66%
≥ 2	{skin ≥ 2} or {fce ≥ 1} or {air ≥ 1} or {frag ≥ 1} or {dye ≥ 1} or {con ≥ 1} or {rec ≥ 1} or {exp ≥ 1}	100%
≥ 3	frag ≥ 2	78%
≥ 3	con ≥ 2	98%
≥ 3	exp ≥ 2	85%
≥ 4	rec ≥ 3	77%

ditional information that may be used to design additional personalized user interface enhancements.

4 Discussion

When comparing the outcome of our original procedure with the outcome of our procedure using DRSA, results are quite interesting. First, examination of the initial phase, the consumer clustering phase, was conducted. When referring to Tables 3 and 6, the reducts obtained differ greatly. When using CRSA, which included the additional step of discretization of features, smaller reducts were obtained, i.e. the majority of the top ten reducts were comprised of only *two* of the 16 features. When using DRSA, which omitted the discretization of features given the dominance-based characteristic of DRSA, five reducts were found, the majority of which were comprised of at least *four* of the *eight* features. These results may indicate the higher complexity of DRSA reduction criteria in relation to CRSA. However, even though DRSA may be more complex in nature,

its appropriateness for this task is apparent given its unique quality to evaluate dominance-based relations.

The next goal of the evaluation was to compare and contrast associations obtained from the second phase of our first evaluation, the association mining phase using Eclat with results obtained using DRSA. As previously mentioned, results described in Tables 4, 5, 7, and 8 illustrate the uniqueness of each algorithm, although some similarities are visible, e.g. both Eclat and DRSA made visible the distinction of rankings for *recyclable packaging* between respective consumer/product clusters. The results obtained in the first evaluation using Eclat provide a deeper view of preferences in terms of the individual group preferences and products whereas DRSA provides a broader overview of preferences of the entire user group and product set. It is hypothesized that both methods could be incorporated into the design of our procedure.

As such, results obtained from DRSA may also provide a possible third phase to our original procedure. Here, knowledge obtained could be used to incorporate design enhancements in terms of the organization of products and features on the user interface. For example, it may be that consumers may also be interested in products designated to other clusters. Results obtained from DRSA analyses provide an indication of cross-cluster product feature and feature ranking similarities. In this sense, there may exist products in other clusters that consumers may be interested in, regardless of what cluster membership they have been initially assigned to as per our procedure.

5 Conclusion

This paper described an evaluation of DRSA in relation to a procedure we developed that originally used CRSA and Eclat to gather information in support of designing personalized user interfaces for online shopping support tools. Specifically, the two phases of our original procedure were analytically and empirically evaluated using results obtained when using DRSA with our procedure. In the first phase, when comparing CRSA with DRSA, the results indicated that classification accuracy was slightly higher when using CRSA. However, it is hypothesized that DRSA may be more appropriate given its unique consideration of preference-ordered data. In the second phase, we found that our original procedure could be further developed using newly obtained information for designing additional design enhancements with results obtained from DRSA. As such, a possible third phase to our original procedure was discussed, incorporating results obtained from DRSA analyses to highlight products and features that target cross-cluster similarities between product feature values and consumer preferences.

Future work will include a more detailed evaluation, specifically of the second phase of our experiment using only features comprising the reducts versus all features when using Eclat as well as incorporating both the discretized and non-discretized features in the analysis. Future work will also include evaluation of prototype designs using both our original procedure using CRSA-Eclat and the procedure described in this paper using DRSA to determine the best possible approach and procedure for providing personalized user interface designs.

Acknowledgments

References

[1] J. Bazan and M. Szczuka. The Rough Set Exploration System. *Transactions on Rough Sets 3*, pages 37–56, 2005.

[2] C. Borgelt. Implementations of Apriori and Eclat. *In Proc. Workshop on Frequent Item Set Mining Implementations (FIMI)*, 2003.

[3] A. Ceglar and J. F. Roddick. Association Mining. *ACM Computing Surveys*, 38(2), 2006.

[4] S. Greco, B. Matarazzo, and R. Slowinski. A New Rough Set Approach to Multicriteria and Multiattribute Classification. *In Proc. Rough Sets and Current Trends in Computing (RSCTC)*, pages 60–67, 1998.

[5] S. Greco, B. Matarazzo, and R. Slowinski. Rough Sets Theory for Multicriteria Decision Analysis. *European Journal of Operational Research*, pages 1–47, 2001.

[6] S. Greco, B. Matarazzo, and R. Slowinski. Rough Approximations by Dominance Relations. *International Journal of Intelligent Systems*, 17:153–171, 2002.

[7] S. Greco, B. Matarazzo, and R. Slowinski. Rough Set Approach to Customer Satisfaction Analysis. *In Proc. Rough Sets and Current Trends in Computing (RSCTC)*, pages 284–295, 2006.

[8] D. H. Hepting. Ethics and Usability Testing in Computer Science Education. *ACM Special Interest Group on Computer Science Education (SIGCSE)*, 38(2):76–80, 2006.

[9] W. Li, V. C. Arena, N. B. Sussman, and S. Mazumdar. Model Validation Software for Classification Models using Repeated Partitioning: MVREP. *Computer Methods and Programs in Biomedicine*, pages 81–87, 2002.

[10] T. Maciag, D. H. Hepting, and D. Ślęzak. Personalizing User Interfaces for Environmental Decision Support Systems. *In Proc. Rough Sets and Soft Computing in Intelligent Agent and Web Technology*, 2005.

[11] T. Maciag, D. H. Hepting, and D. Ślęzak. Consumer Modelling in Support of Interface Design. *In Proc. IEEE International Conference on Hybrid Information Technology*, 2:153–160, 2006.

[12] T. Maciag, D. H. Hepting, D. Ślęzak, and R. J. Hilderman. Evaluating the Utility of Web-Based Consumer Support Tools Using Rough Sets. *To Appear In Proc. International Conference on Conceptual Structures (ICCS), Rough Set and Data Mining Workshop*, 2007.

[13] T. Maciag, D. H. Hepting, D. Ślęzak, and R. J. Hilderman. Mining Associations for Interface Design. *In Proc. International Conference on Rough Sets and Knowledge Technology (RSKT), Joint Rough Set Symposium (JRS)*, pages 109–117, 2007.

[14] B. Mobasher, R. Cooley, and J. Srivastava. Automatic Personalization Based on Web Usage Mining. *Communications of the ACM*, 43(8):142–151, 2000.

[15] Z. Pawlak. *Rough Sets, Theoretical Aspects of Reasoning About Data*. Kluwer Academic Publishers, 1991.

[16] J. Raskin. *The Humane Interface: New Directions for Designing Interactive Systems*. Addison-Wesley, 2000.

[17] H. A. Simon. A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, 69:99–118, 1955.

[18] R. Slowinski, S. Greco, and B. Matarazzo. *Rough Set Based Decision Support*, chapter 16. Springer-Verlag, 2005.

[19] U. S. United States Environmental Protection Agency. Cleaning Products Pilot Project Fact Sheet. [http://www.epa.gov/oppt/epp/pubs/cleanfct.pdf] PDF, 1997.

[20] M. J. Zaki. Scalable Algorithms for Association Mining. *In Proc. IEEE Transactions on Knowledge and Data Engineering*, 12(3), 2000.